

RELACION ENTRE EL ESTRATO Y LA UBICACION DE LUGARES EN BOGOTÁ

David Mora

Octubre, 2020

1. Introducción

1.1 Contexto

Mi ciudad natal, Bogotá, cuenta con una población cercana a 8 millones de habitantes y está dividida geográficamente en 20 localidades cada una con características diferentes a la vez cada localidad está dividida geográficamente por barrios para un total de 1922 barrios. Además de esta división geográfica, Bogotá tiene una división por estratos, cada zona de la ciudad está clasificada dependiendo de las características de las viviendas, el entorno urbano de la zona y el contexto urbanístico. Es así como la ciudad se subdivide en 6 estratos socio-económicos, siendo el 1 el más bajo y el 6 el más alto. Esto con el fin de identificar zonas de acción y distribuir el costo de los servicios públicos; en donde los estratos más altos subvencionan a los más bajos y estos a la vez pueden acceder a beneficios educativos o de salud dada la estratificación. Esto le ha permitido a la ciudad identificar rápidamente sectores vulnerables y es así como, entre otras cosas, ha logrado garantizarle gratuitamente a los estratos 1 y 2 el consumo mínimo vital de agua (primera y única ciudad de Colombia). Es necesario añadir que con base a la importante cifra de patrimonio arquitectónico, cultural, e histórico que hay en la ciudad de Bogotá, se creó el estrato 1 patrimonial, que brinda de los beneficios mencionados anteriormente a los poseedores de esos predios.

Según The Economist, en el plano económico, Bogotá se destaca por su fortaleza económica asociada al tamaño de su producción, las facilidades para crear empresas y hacer negocios, la madurez financiera, la atracción de empresas globales y la calidad de su capital humano. Es el principal mercado de Colombia y de la Región Andina, y el primer destino de la inversión extranjera directa que llega a Colombia (70 %). Tiene el mayor PIB nominal y per cápita del país, aportando la mayor parte al total nacional (24,5 %), y es la sexta ciudad por tamaño del PIB en Latinoamérica (de unos USD 106 678 millones) con un PIB per cápita nominal de USD 11.594 y un PIB PPA per cápita de USD 20.120, igualmente es la plataforma empresarial más grande de Colombia en donde ocurren la mayoría de los emprendimientos de alto impacto. Las anteriores razones hacen que Bogotá sea objetivo empresarial de marcas reconocidas a nivel mundial.

Bogotá es considerada una ciudad multicultural debido a que ha recibido habitantes de todas las regiones del país, ya sea porque buscan mejores oportunidades para sus negocios, mayor oferta educativa o por el conflicto armado que ha golpeado desde los años 50 las regiones rurales del territorio colombiano. Las personas que llegan a Bogotá por razones de desplazamiento forzado son personas que en su mayoría no tienen los suficientes recursos para subsistir en la ciudad por lo que se hace imperante vivir en los barrios de estratos más bajos donde el pago de servicios públicos es el más bajo posible pero las condiciones de vida son más difíciles.

1.2 Problema e interesados

1.2.1 Problema 1: Marcas Internacionales en Bogotá

La decisión de donde poner un punto de ventas depende de muchas variables entre las cuales podemos nombrar el público objetivo, precio del alquiler, precio de los servicios públicos, percepción de seguridad, de carácter económico del barrio entre otros aspectos. Estos aspectos están muy relacionados en Bogotá con el estrato de la ubicación del lugar, por ejemplo en una ubicación en estrato alto el alquiler y servicio serán más caros pero se tendrá acceso a clientes con mejor capacidad económica. A toda empresa que este iniciando sus actividades en Bogotá le interesa saber en dónde ubicar sus puntos de ventas o de trabajo.

1.2.2 Problema 2: Ubicación de los lugares en Bogotá según la opinión de los usuarios en contraste con su estrato

Para la **alcaldía de Bogotá** y las **instituciones que monitorean la calidad de vida** de sus habitantes tal como la secretaria de habitad, ministerio de salud y cámara de comercio de Bogotá entre otros, es importante conocer la opinión de sus ciudadanos e identificar si hay una tendencia en la ubicación de los lugares con mejor calidad en servicio y si estos lugares tienen la variedad de servicios adecuados para suplir las necesidades que incrementan la calidad de vida de los ciudadanos.

Los indicadores de calidad de vida que se tienen en cuenta para dar el índice de calidad de vida según la firma Mercer (<https://mobilityexchange.mercer.com/quality-of-living-reports>) que ubico a Bogotá en la posición 128 entre mas de 450 ciudades son los siguientes: bienes de consumo, ambiente económico, alojamiento, consideraciones médicas y de salud, entorno natural, entorno político y social, servicios públicos y transporte, recreación, escuelas y educación, entorno sociocultural. De los anteriores indicadores considero que se relacionan con la información que se puede obtener de FourSquare son los siguientes bienes de consumo, ambiente económico, ambiente socio cultural y recreación.

2. Adquisición y limpieza de los datos

Para los dos problemas planteados se requieren la siguiente información: mapa de Bogotá, estrato de cada cuadra de Bogotá, lugares que hay en Bogotá y para el primer problema se necesitaría adicional lo anterior, un listado de marcas reconocidas a nivel internacional.

2.1 Mapa de Bogotá

Se obtuvo de la página <https://gist.github.com/john-guerra/727e8992e9599b9d9f1dbf4c4c8e479e>. Este esta descrito mediante un archivo GeoJson que contiene el contorno de todos los departamentos de Colombia descritos mediante polígonos, se seleccionó solo el polígono de Bogotá.

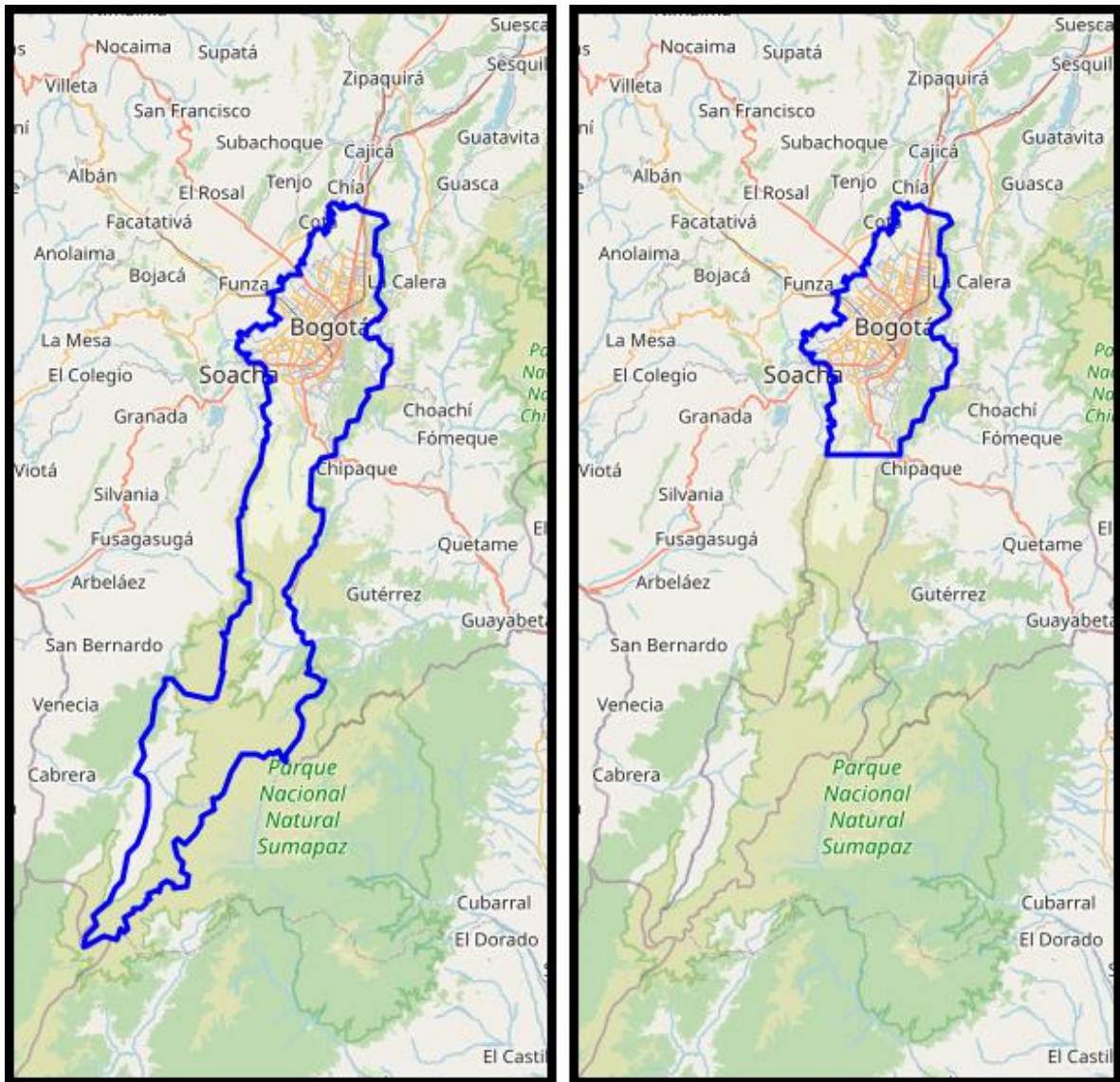


Figura 1. Mapa del contorno de Bogotá, a la izquierda el contorno completo de Bogotá, a la derecha el contorno que se usó, se recortó la zona sur que corresponde a una zona rural.

En la figura 1 a la izquierda se puede observar el contorno de Bogotá. Dado que más del 60% de Bogotá es zona rural y nuestro análisis se enfoca en el casco urbano se recortó el mapa en la latitud 4.459243 de tal forma que se conservó una zona rural menor a la zona urbana, el este mapa reducido de Bogotá se puede observar en la figura 2 a la derecha.

2.2 Estrato de cada cuadra de Bogotá

Esta información se encuentra en un archivo GeoJson obtenido de la página de datos abierto de Bogotá, en particular en el apartado <https://datosabiertos.bogota.gov.co/dataset/manzana-estratificacion-bogota-d-c>. Este archivo contiene cada cuadra de Bogotá consignada en polígonos que están asociados al estrato de cada respectiva cuadra, en la figura 2 se puede observar algunas cuadras, una de ellas con su estrato.



Figura 2. Mapa con los polígonos de cada cuadra de Bogotá y su información incluido el estrato.

2.3 Lugares de Bogotá

Se formó un dataframe con la información de 7488 lugares de Bogotá que se obtuvo mediante peticiones regulares y premium a Four Square, el dataframe contiene las siguientes columnas: nombre, categoría, latitud, longitud, id, numero de punto, Estrato del lugar, calificación, numero de calificadoros, cantidad de fotos, cantidad de opiniones.

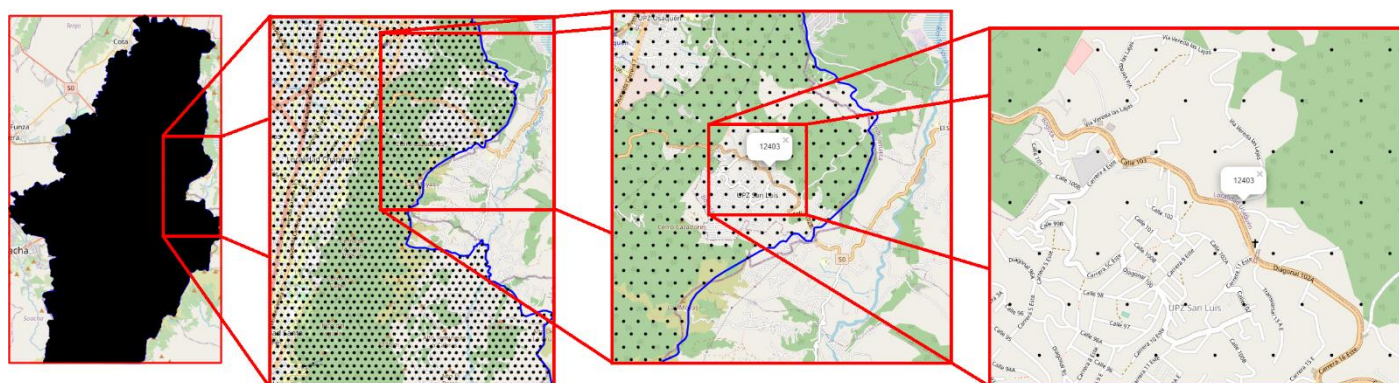


Figura 3. Rejilla de puntos sobre los que se hicieron solicitud de lugares a Four Square en Bogotá.

Las peticiones se hicieron en ubicaciones dadas por una rejilla triangular (se escogió de esta forma para optimizar el número de ubicaciones solicitadas) sobre el mapa del contorno de Bogotá con un radio de 129m. En la figura 3 se puede apreciar la rejilla dispuesta sobre el área dentro del contorno del Bogotá y en la figura 4 se puede observar una zona en la que se muestran los lugares obtenidos en los puntos de la rejilla, cada punto esta numerado y cada lugar encontrado está asociado al punto nombrado en la columna NumPunto. El radio y la distancia entre puntos en la rejilla se configuraron de tal forma que se obtuvieran la mayor cantidad de lugares posibles.

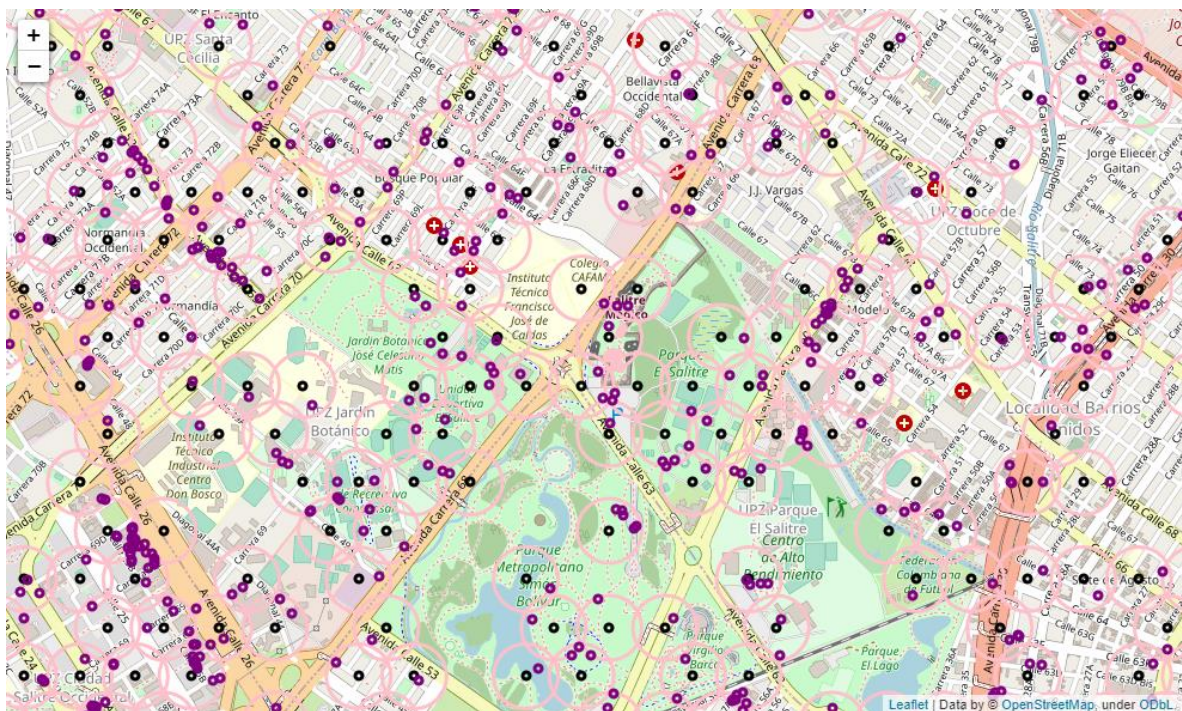


Figura 4. Mapa de una región de Bogotá, en morado los lugares encontrados en una circunferencia color rosa centrada en los puntos de la rejilla en color negro.

Usando la ubicación de los lugares obtenidos y la información del estrato de cada cuadra (numeral 2.2) se estableció el estrato en el que está ubicado el lugar, en algunos casos las ubicaciones no estaban bien, quedando sobre avenidas o aceras, en ese caso se asignó el estrato de la cuadra más cercana. Se hizo una limpieza de lugares que presentarían ubicaciones fuera de lo común, por ejemplo tiendas de implementos deportivos o un mc'donnals en medio de un bosque en la periferia de la ciudad. En la figura N se pueden observar los títulos de las columnas y las cinco primeras y últimas filas del dataframe obtenido.

	name	categories	lat	lng	id	NumPunto	Estrato	CRat	Rat	CFot	Copi
0	El Fogón Del Llano	Cajun / Creole Restaurant	4.468930	-74.127996	4f4a782de4b0cccedd61818b	134	2	0.0	0.0	0.0	0.0
1	Plaza de Usme	Plaza	4.469970	-74.126363	4f4a3862e4b0eddb52dd97bb	135	2	0.0	0.0	12.0	0.0
2	Juan Valdez	Café	4.477763	-74.122131	537260f7498e7c622a9bf03a	245	1	0.0	0.0	0.0	0.0
3	agroparque los soches	Field	4.483056	-74.091950	51ebe330498e9a833ac50	367	1	0.0	0.0	3.0	0.0
4	Lacteos Santa Monica	Health Food Store	4.490626	-74.148808	5bd1f68e12c8f0002cc8b39c	445	2	0.0	0.0	0.0	0.0
...
7483	TÍPICA sopas y picadas	Steakhouse	4.815070	-74.031742	53d41873498e67ffa980e88d	14903	6	12.0	6.4	13.0	5.0
7484	Club Chalaneria Guadalupe	Stables	4.817814	-74.069090	563e6a79498e41deb3b8dab	14926	6	0.0	0.0	0.0	0.0
7485	Al Limite Paintball	Paintball Field	4.818875	-74.032603	4ca77d9ba8e08cfa63508594	14944	6	37.0	8.2	51.0	13.0
7486	Refugio Hípica - Club Guaymaral	Stables	4.820947	-74.039281	503913f2e4b0cc82fd815e42	14978	6	0.0	0.0	11.0	1.0
7487	Bogotá Tennis Club	Tennis Court	4.820655	-74.036786	4bc217ed74a9a5935cb9d2f6	14979	6	55.0	8.0	36.0	8.0

7488 rows x 11 columns

Figura 5. Dataframe obtenido luego de las solicitudes a Four Square.

Inspeccionando los valores de la columna Rat se observó que en la mayoría de casos en que CRat es cero, Rat esta también lo era, esto se debe a que cuando no se ha dado ninguna calificación FourSquare no da ningún valor, en ese caso se decidió colocar el valor en cero. Se observó que 14

casos a pesar de no haber calificaciones se daba un valor en la columna Rat, esto se puede deber a que FourSquare importa dicha información de otras fuentes en las que no se tiene la cantidad de calificadores o simplemente es información que se perdió por alguna razón desconocida.

2.4 Listado de marcas reconocidas a nivel internacional

Se hizo scraping de dos fuentes: GLOBAL 2000 The World's Largest Public Companies (<https://www.forbes.com/global2000/>) y About World Top Companies Value (<https://www.value.today/world/world-top-1000-companies>), se obtuvieron las 2000 y 1000 nombres de marcas con mayores ganancias a nivel mundial respectivamente. Por lo menos 15 marcas que reconozco como internacionales no aparecieron en los listados obtenidos en el scraping, por ejemplo: KFC, Buffalo Wings, Burger King, Zara, Lego, Reebok, Subway, por mencionar algunos por lo que se intercambiaron con marcas que se comprobó previamente no se encuentran en Bogotá.

En ambas listas fue necesario eliminar algunos acrónimos y palabras como S.A., LTD., .com, PLC, inc, incorporated, limited, group, company, con el fin de evitar coincidencias erradas en el proceso.

Recurriendo al dataframe conformado en la sección 2.3 se seleccionaron los nombres de las marcas que coincidían en alguno de los dos listados, se presentaron algunas coincidencias con marcas no internacionales debido al nombre de unas marcas pero estas fueron limpiadas, ejemplo de esto la palabra Santander a nivel internacional es reconocida como un banco español y en Colombia es el apellido de un héroe de las guerras de independencia en consecuencia es el nombre de un departamento de Colombia por lo que es muy usado en nombres de restaurantes de comida típica de ese departamento. Se llegó a un dataframe con 334 lugares que pertenecen a marcas internacionales en Bogotá y la información de cada lugar.

3. Metodología y resultados

3.1 Problema 1: Marcas Internacionales en Bogotá

A partir del dataframe obtenido en 2.4 se determinó que los 334 lugares están compuestos por 27 marcas y 29 categorías diferentes. Se hizo una agrupación por estrato para determinar la cantidad de lugares que hay en cada estrato, estos valores se pueden apreciar en la figura 6. Se puede observar que el mayor porcentaje de lugares internacionales están ubicados en los estratos 3 y 4 (68.8%), seguidos de los estratos 5 y 6 (27.6%), y en menor porcentajes en estratos 1 y 2 (3.6%) donde es de resaltar que en el estrato 1 no se encontró ninguna marca internacional.

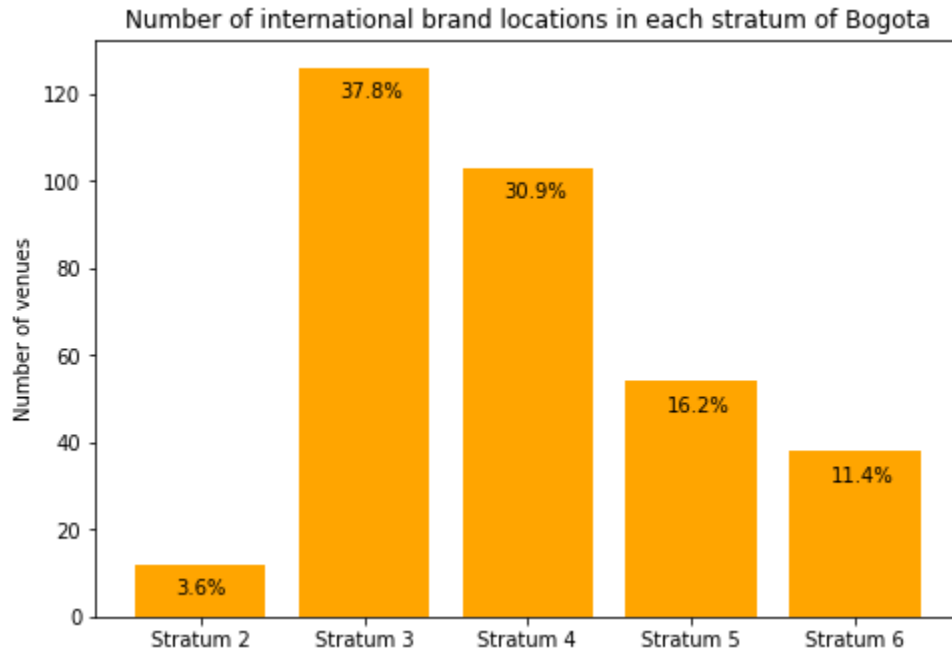


Figura 6. Diagrama de barras por estratos de los lugares de marcas internacionales en Bogotá.

Al observar el histograma de número de lugares que tiene cada marca con 20 barras en la figura 7, se puede ver que cerca de la mitad de marcas tienen entre 1 y 5 lugares, lo cual se vio también al hacer la estadística descriptiva de la agrupación por marcas.

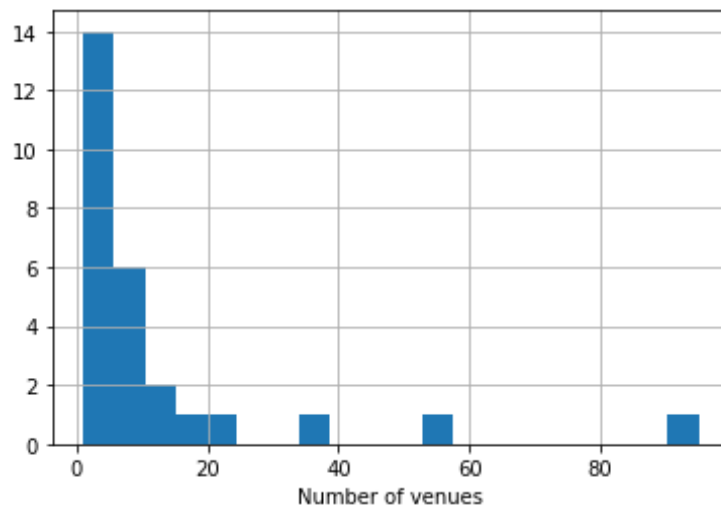


Figura 7. Histograma del número de lugares de marcas internacionales en Bogotá

Se hizo una agrupación por marcas en la que se encontró que hay marcas con una única ubicación hasta marcas con 95 ubicaciones, se encontró que por lo menos el 25% de marcas tiene una sola ubicación, el 50% tiene 4 o menos ubicaciones y el 75% tiene 12.5 ubicaciones o menos.

Dado que se busca saber en qué estrato prefieren ubicarse las marcas internacionales se hizo también una agrupación por marca y se obtuvo la media de cada estrato, de esta forma marcas

como Subway, Dunkin Donuts o McDonald's estarían representados por un solo lugar y estrato en el que deciden ubicarse. En la figura 8 se puede observar la distribución de las marcas internacionales por estratos cuando se toma el promedio de estrato por marca y en esta se puede ver que las marcas internacionales no tienen ningún punto en el estrato mas bajo de la ciudad (estrato 1), prefieren ubicarse en el estrato 4 y hay más marcas que se ubican en estrato 5 o 6 que en los estratos 2 o 3.

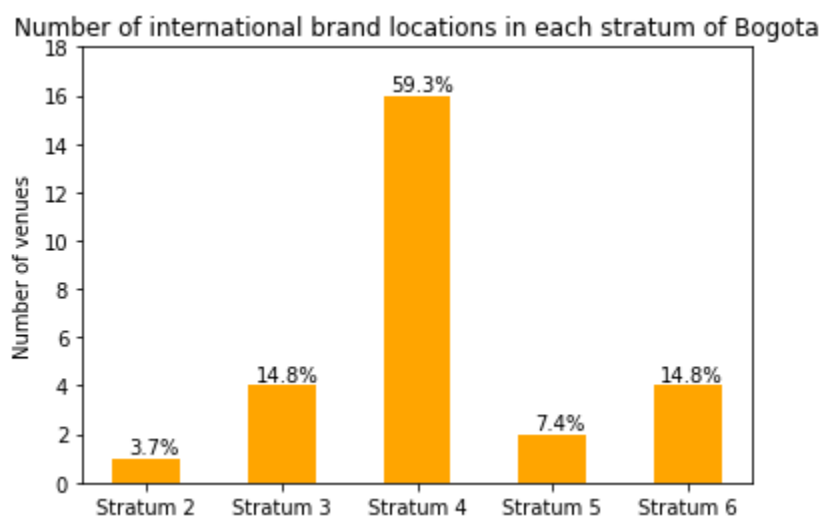


Figura 8. Numero de marcas internacionales en cada estrato.

Luego se procedió a hacer un análisis por cuartiles en la cantidad de lugares que tiene cada marca.

3.1.1 Primer Cuartil

En la figura 9 se puede observar la distribución por estratos de marcas que tienen solo un lugar en la ciudad, esto corresponden a 10 marcas es decir el 37% de las marcas internacionales.

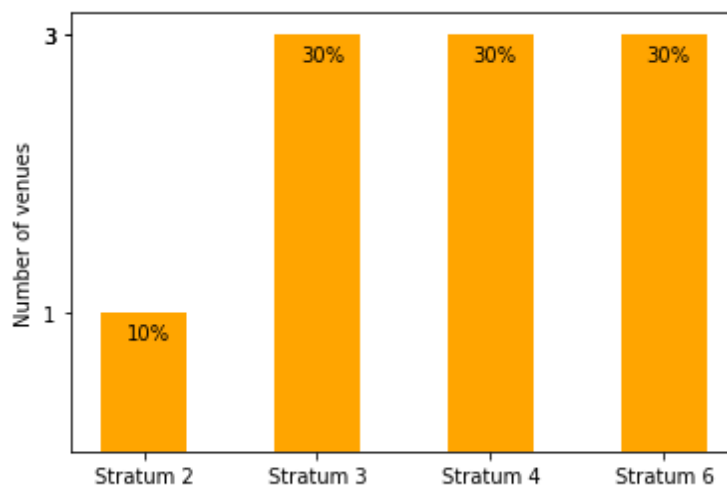


Figura 9. Numero de lugares por estrato cuando las marcas internacionales tienen un solo lugar en Bogotá.

La anterior grafica nos muestra que cuando las marcas internacionales tienen un solo punto en la ciudad, prefieren ubicarse en estratos medio y alto, de esta forma tienen la posibilidad de atraer clientes con mayores ingresos.

3.1.2 Segundo y Tercer Cuartil

En la figura 10 se puede observar la distribución por estratos de marcas que tienen entre cuatro y doce lugares en la ciudad estos corresponden a 8 marcas de las 27 marcas internacionales encontradas en Bogotá, estas 8 marcas acumulan un total de 59 lugares.

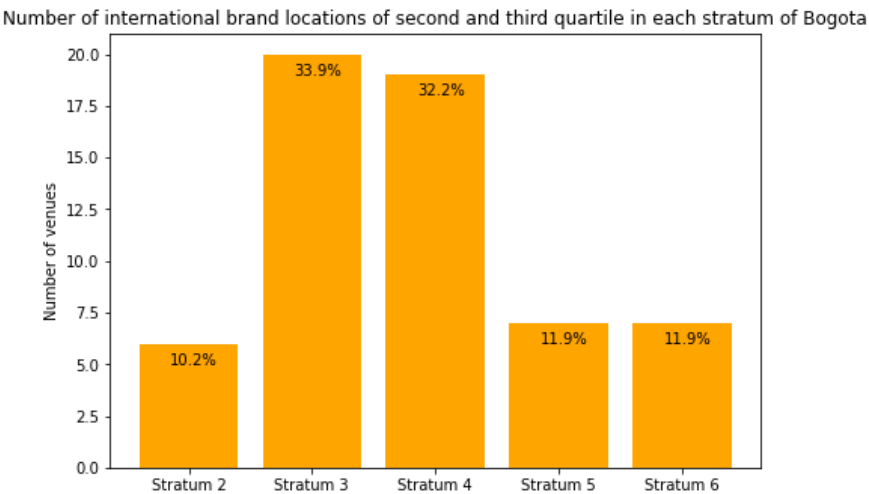


Figura 10. Numero de lugares por estrato cuando las marcas internacionales tienen entre 4 y 12 lugares en Bogotá.

En la figura 10 podemos ver que las marcas que tienen entre cuatro y doce lugares en la ciudad tienen a ubicarse principalmente en estratos medio (3 y 4) y en estratos altos (23.8% en estratos 5 y 6)

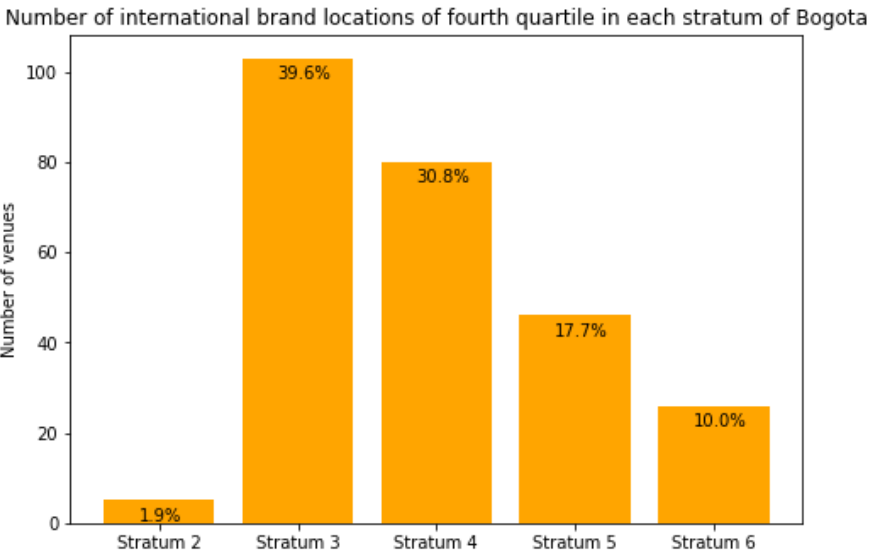


Figura 11. Numero de lugares por estrato cuando las marcas internacionales tienen más de 12 lugares en Bogotá.

3.1.3 Cuarto Cuartil

Aquí se puede observar el comportamiento de las marcas internacionales que más lugares tienen en Bogotá. El cuartil cuarto está compuesto por las marcas que tienen más de 13 ubicaciones, estas corresponden a 7 marcas que tienen en total 260 lugares, en la figura 11 se puede observar el número de lugares que tienen en cada estrato las marcas con 13 o más lugares en Bogotá, se puede observar que la mayoría de marcas en este cuartil tienen sus locales ubicados principalmente en los estratos medio (estrato 3) y medio alto (estrato 4), 39.6% y 30.8% respectivamente y en menor medida en estratos altos, estratos 5 y 6 (27% en sumando ambos) y una minoría en el estrato 2 (1.9%).

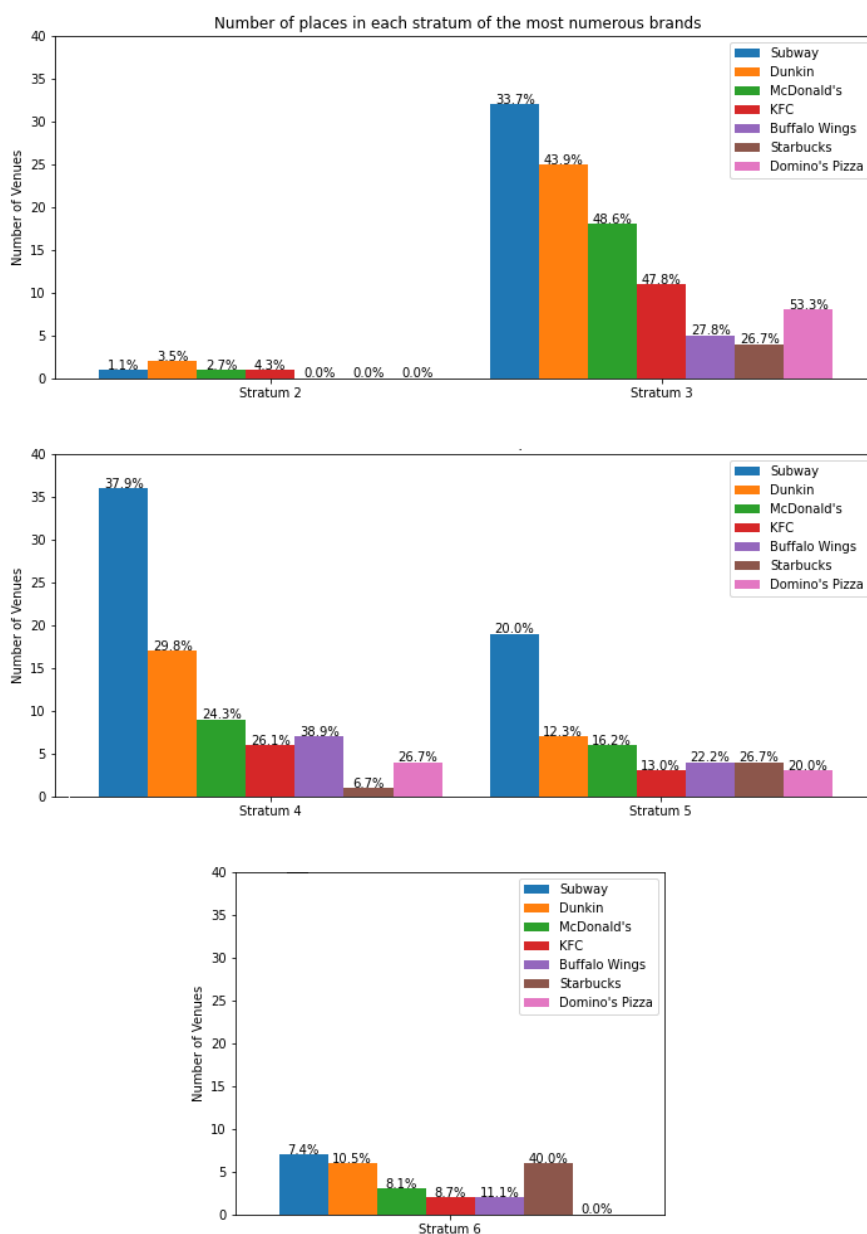


Figura 12. Numero de lugares en cada estrato de las marcas internacionales con más de 12 lugares en Bogotá

En la figura 12 se puede observar el porcentaje de lugares que tienen en cada estrato cada una de las marcas internacionales con más de 12.5 lugares en Bogotá. Se puede observar que la mayoría de marcas internacionales con más de 13 lugares en Bogotá prefieren ubicarse en estrato 3 y disminuyen gradualmente el número de lugares que tienen en cada estrato a medida que este es más alto, solo Subway y Buffalo Wings tienen más puntos en estrato 4 que en estrato 3 y Starbucks es el único que tiene más puntos de venta en el estrato 6 que en el resto de estratos.

3.2 Problema 2: Ubicación de los lugares en Bogotá según la opinión de los usuarios en contraste con su estrato

Se inició haciendo una agrupación por estratos del dataframe obtenido en la sección 2.3. Se puede ver en la figura 13 la distribución de lugares por estratos en Bogotá. Al hacer una agrupación por categorías de encontró que hay 357 categorías diferentes, las 5 primeras corresponden a sitios de ventas de comida (restaurante, punto de hamburguesas, panadería, restaurante de comida rápida, lugar de pizza).

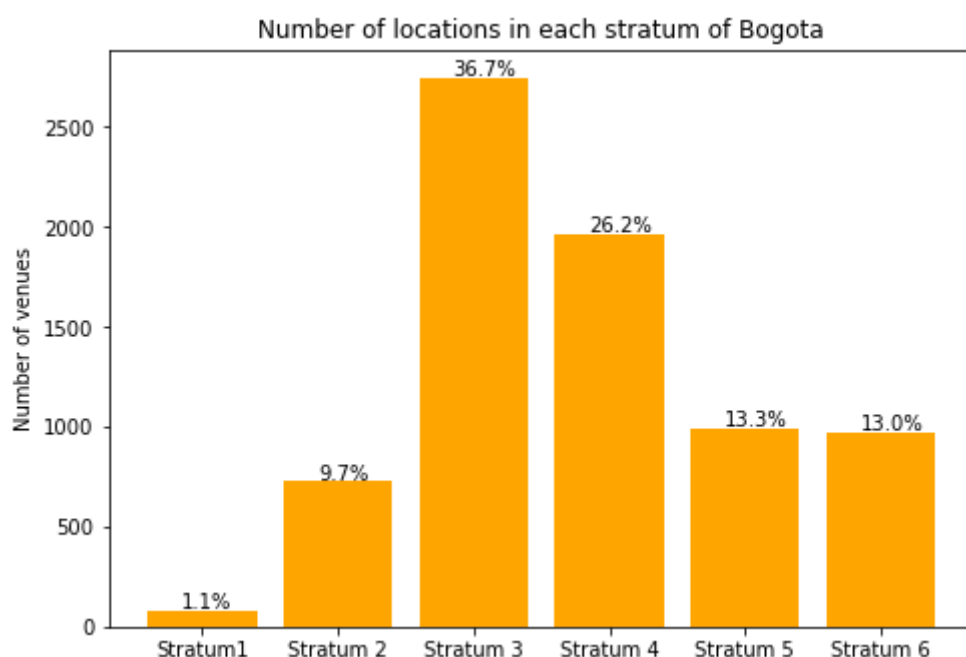


Figura 13. Diagrama de barras del número de lugares en cada estrato de todos los lugares encontrados en Bogotá

Recordemos que del dataframe obtenido en la sección 2.3 las últimas 4 columnas corresponden a la imagen que tienen los usuarios sobre el lugar, entiéndase por imagen a que tanto están dispuestos los usuarios por ejemplo a recomendarlo, a opinar sobre él, a dar una calificación independiente de que sea alta o baja o a subir una foto del lugar; estas columnas tienen las siguientes abreviaciones: cantidad de calificaciones (CRat), calificación promedio (Rat), cantidad de fotos (CFot) y cantidad de opiniones (COpi). Para este trabajo por simplicidad se tendrá en cuenta que una opinión dada, una foto subida o una calificación aportada tienen el mismo peso. En ese orden de ideas las columnas CRat, CFot y COpi pueden tener valores de cero en adelante y la columna Rat solo puede tener valores entre 0 y 10.

Se hizo histogramas y diagrama de caja para cada columna de calificación de imagen, estas se puede observar en la figura 14. Se puede apreciar que en todos los casos las barras de que cubren el primer intervalo tienen la mayoría de datos, esto implica que hay muchos lugares que tienen cantidades en cero o cercanos a cero, además de eso se puede ver que en las columnas CRat, CFot y COpi los ejes horizontales aparecen con el orden de miles pero con barras de altura mínima, además en el diagrama de cajas se puede apreciar que hay valores atípicos del orden de miles en esas mismas columnas, esto implica que hay unos pocos lugares que tienen muchas calificaciones, fotos y opiniones. Teniendo esto en cuenta se decidió organizar nuestro análisis de calificaciones en tres categorías: cuando los tres valores son ceros, cuando hay 1 o 2 ceros y cuando no hay ceros en las columnas CRat, CFot y COpi.

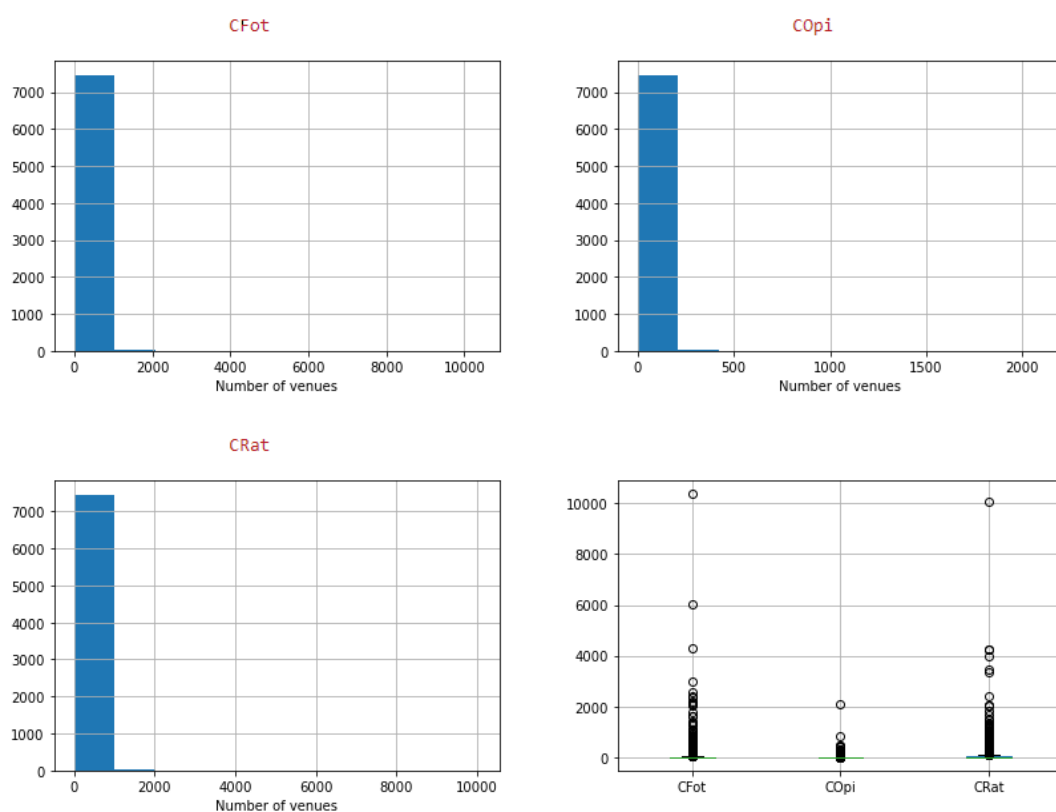


Figura 14. Histogramas y diagrama de caja de las columnas de apreciación de los usuarios de los lugares de Bogotá

3.2.1 Todos los valores son ceros

Se seleccionaron las filas del dataframe que tenían todos los elementos con valor cero en las columnas de apreciación de los lugares de Bogotá, quedando solo 750 lugares (10.02% del total de lugares), luego se agruparon por estratos y se obtuvo la distribución mostrada en la gráfica 15.

Dada la condición de calificaciones de estos lugares se considerara que son los peor calificados ya que ningún usuario o los mismos propietarios se han tomado la tarea de calificar el lugar. Se puede apreciar que estos lugares se encuentran principalmente en el estrato 3, seguido del estrato 2 y el estrato 4 y en menor medida en los estratos 1, 5 y 6, esto nos indica que los lugares que no tienen ningún tipo de calificación se encuentran en mayor medida en los estratos medio y bajo.

No se aplicó clustering a este grupo de datos ya que no hay forma de distinguir su comportamiento en cuanto a sus calificaciones.

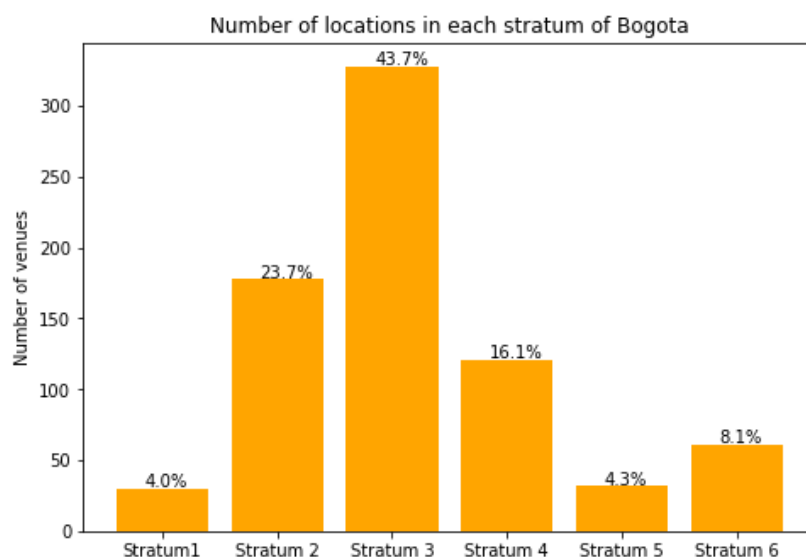


Figura 15. Diagrama de barras por estrato de los lugares que tienen las columnas de apreciación en cero.

3.2.2 Hay 1, 2 o 3 valores en cero

Se seleccionaron las filas del dataframe que tenían 1, 2 o 3 ceros en las columnas de apreciación de los lugares de Bogotá, luego se agruparon por estratos y se obtuvo la distribución mostrada en la gráfica 16. Dicho dataframe quedo con un total de 2424 registros (32.37% del total de lugares). Se encontraron 414 lugares con un cero (5.53% del total), 902 lugares con 2 ceros (12.05% del total) y 1108 lugares con 3 ceros (14.80% del total).

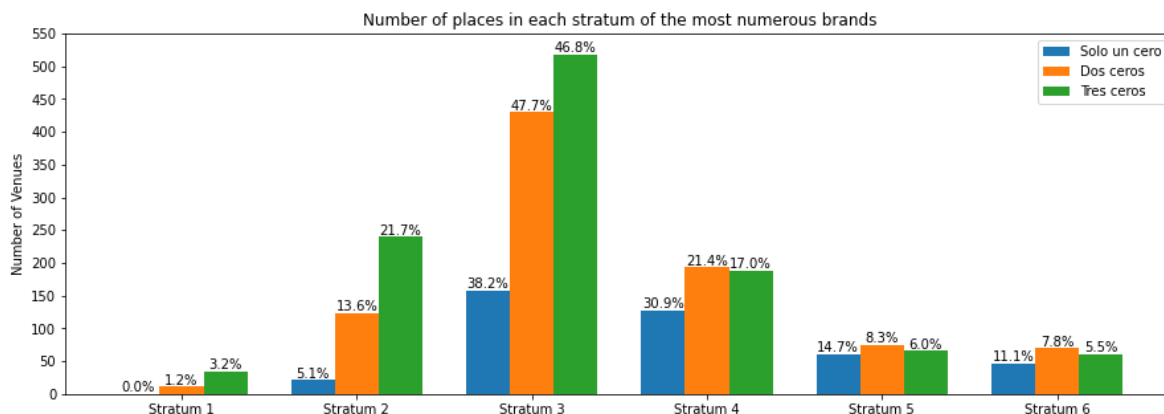


Figura 16. Diagrama de barras por estrato de los lugares que tienen las columnas de apreciación en cero.

En figura 17 se muestra la distribución de la sumatoria de los valores que hay en cada estrato, se puede observar que en total predomina en primer posición el estrato 3 y en segundo lugar el estrato 4 y en tercer lugar el estrato 2 y en menor medida los estratos 1, 5 y 6.

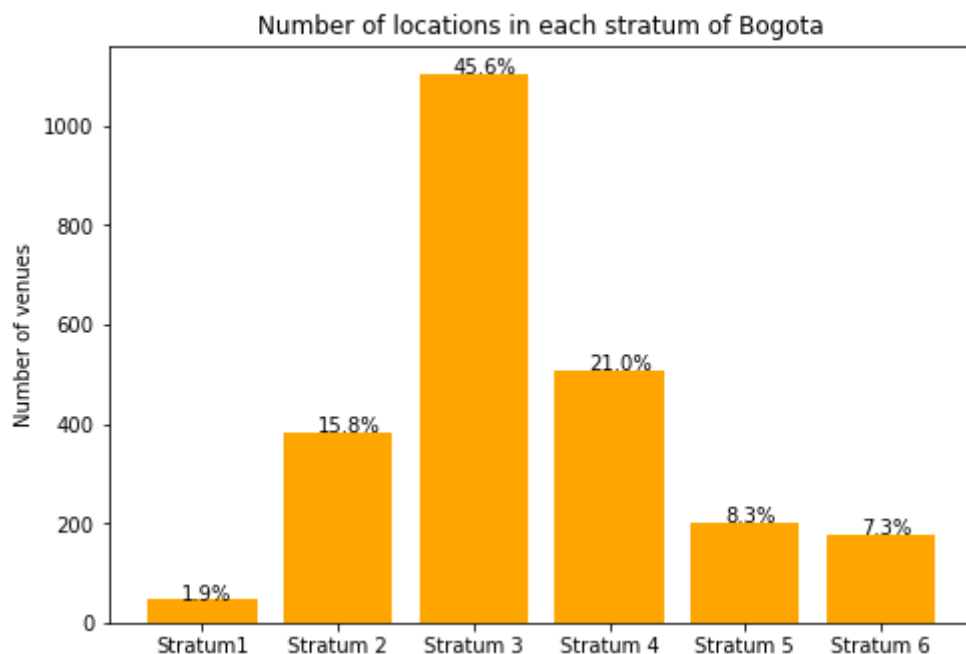


Figura 17. Diagrama de barras por estrato de los lugares que tienen en las columnas de apreciación 1, 2 o 3 ceros.

Al observar los valores máximos en las columnas de valoración se encontró que hay lugares con una cantidad alta de opiniones (35), fotos (306) y calificaciones (69) en comparación con las medias respectivas. También es de notar que el estrato promedio fue 3.39, al observar la descripción estadística de las columnas de apreciación en la figura 18, podemos ver desde que valores en adelante son considerados valores atípicos. La fila SumC corresponde a la estadística de la suma de las columnas CRat, COpi y CFot, las columnas RanIQ es el rango intercuartílico, AtipLevelInf el valor inferior del intervalo de valores no atípicos, AtipLevelSup el valor superior del intervalo de valores no atípicos.

	count	mean	std	min	25%	50%	75%	max	RanIQ	AtipLevelInf	AtipLevelSup
CRat	2424.0	2.337046	5.881679	0.0	0.0	0.0	0.0	69.0	0.0	0.0	0.0
COpi	2424.0	0.851073	1.704699	0.0	0.0	0.0	1.0	35.0	1.0	-1.5	2.5
CFot	2424.0	3.580858	9.931714	0.0	1.0	2.0	4.0	306.0	3.0	-3.5	8.5
SumC	2424.0	6.768977	12.266753	0.0	2.0	4.0	8.0	315.0	6.0	-7.0	17.0

Figura 18. Descripción estadística columnas de apreciación de los lugares que 1,2 o 3 ceros

3.2.3 No hay ningún valor en cero

Se seleccionaron las filas del dataframe que no tenían valores en cero en las columnas de apreciación de los lugares de Bogotá, de esta forma quedan 4313 registros (57.61% del total de lugares), luego se agruparon por estratos y se obtuvo la distribución mostrada en la gráfica 19.

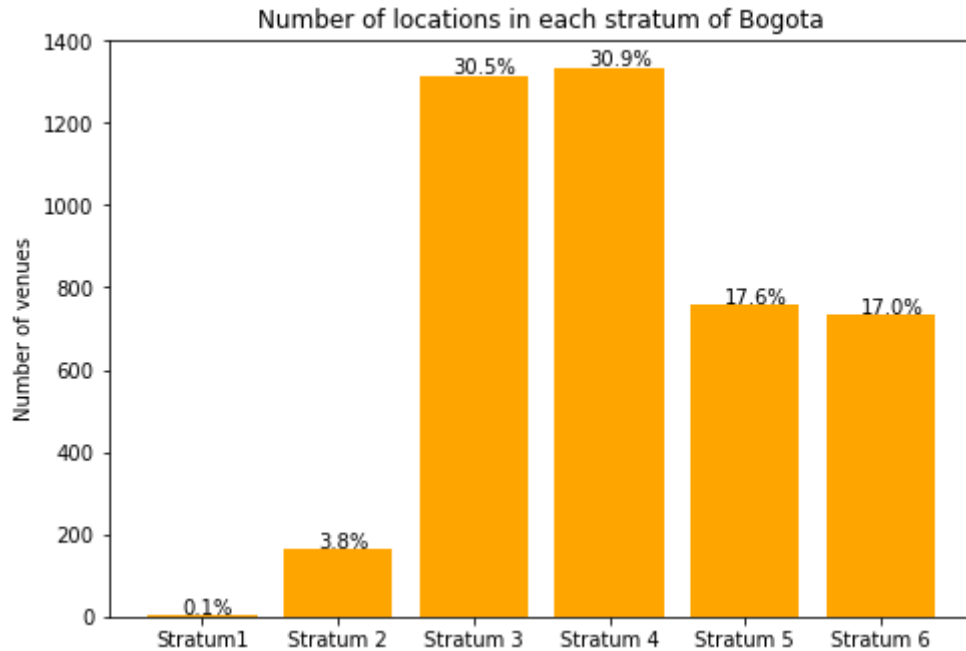


Figura 19. Diagrama de barras por estrato de los lugares que no tienen ceros en las columnas de apreciación.

Se encontró que había lugares cuyos valores eran muy altos en comparación con los valores promedio, opiniones (2105), fotos (10368) y calificaciones (1050). También es de notar que el estrato promedio fue 4.13, al observar la descripción estadística de las columnas de apreciación en la figura 20 podemos ver desde que valores en adelante son considerados valores atípicos. La fila SumC corresponde a la estadística de la suma de las columnas CRat, COpi y CFot, las columna RanIQ es el rango intercuartilico, AtipLevelInf el valor inferior del intervalo de valores no atípicos, AtipLevelSup el valor superior del intervalo de valores no atípicos.

	count	mean	std	min	25%	50%	75%	max	RanIQ	AtipLevelInf	AtipLevelSup
CRat	4314.0	85.280482	254.912375	1.0	13.0	25.0	68.0	10050.0	55.0	-69.5	150.5
COpi	4314.0	17.537552	46.885198	1.0	3.0	6.0	16.0	2105.0	13.0	-16.5	35.5
CFot	4314.0	54.800649	244.082387	1.0	6.0	13.0	36.0	10368.0	30.0	-39.0	81.0
SumC	4314.0	157.618683	526.173577	4.0	24.0	46.0	123.0	22523.0	99.0	-124.5	271.5

Figura 20. Descripción estadística columnas de apreciación de los lugares que no tienen ceros

3.2.4 Clasificación de lugares y Clustering

Se clasificaron los lugares de Bogotá en base a las columnas de apreciación dadas por los usuarios, en la categoría más baja se encuentran los lugares que tienen cero en todas las columnas de apreciación. Se decidió unificar los dataframes de 1, 2 o 3 ceros y el que no tenía ningún cero en las columnas de apreciación debido a que en ambos dataframes se encontraban valores similares en la

columna de suma SumC, esto implica que hay lugares que podríamos clasificar en la misma categoría independiente de si el lugar está en el dataframe de 1, 2 o 3 ceros o el dataframe que no tiene ningún valor en cero. Al unificar estos dataframes se obtuvo uno con un total de 6738 registros. En la figura 21 podemos ver la estadística descriptiva de dicho dataframe, la fila SumC corresponde a la estadística de la suma de las columnas CRat, COpi y CFot, las columna RanIQ es el rango intercuartilico, AtipLevelInf el valor inferior del intervalo de valores no atípicos, AtipLevelSup el valor superior del intervalo de valores no atípicos.

	count	mean	std	min	25%	50%	75%	max	RanIQ	AtipLevelInf	AtipLevelSup
CRat	6738.0	55.441526	207.839831	0.0	0.0	13.0	37.0	10050.0	37.0	-55.5	92.5
COpi	6738.0	11.534580	38.372908	0.0	1.0	3.0	9.0	2105.0	8.0	-11.0	21.0
CFot	6738.0	36.374295	196.927087	0.0	2.0	6.0	20.0	10368.0	18.0	-25.0	47.0
SumC	6738.0	103.350401	427.246973	0.0	6.0	23.0	67.0	22523.0	61.0	-85.5	158.5

Figura 21. Descripción estadística columnas de apreciación de los lugares que no tienen todos ceros.

Debido a que hay valores muy grandes en cada parámetro de apreciación de los usuarios se decidió dividir la clasificación de los lugares en aquellos que tenían una cantidad extremadamente alta en comparación con los demás lugares (los lugares con cantidades muy altas quedan aislados al utilizar el algoritmo de clustering, se observaron clústeres solamente con entre 1 a 10 lugares). Se tomó como criterio de segmentación del dataframe la suma de las columnas CRat, COpi y CFot. Los lugares con 159 o más en suma son fueron considerados atípicos y se consideran en la clasificación más alta. La cantidad de lugares en esta clasificación fue de 868 y al revisar algunos de estos lugares se encontró el estadio de futbol Nemesio Camacho (Estadio principal y más grande de Bogotá), el aeropuerto El Dorado (el aeropuerto más grande de Colombia), centros comerciales, restaurantes de gran fama y lugares abiertos a todo el público como la ciclo vía, estaciones de transporte público o parques públicos.

Con el dataframe resultante de excluir los valores atípicos se aplicó el algoritmo de machine learning no supervisado de clustering desde 3 a 7 clústeres, teniendo como entrada las columnas CRat, Rat, CFot y COpi. Para cada número de clústeres aplicados se obtuvieron las medias de la suma de las columnas de apreciación de cada clúster y se calculó su desviación estándar de las medias con el fin de encontrar el valor más alto, esto con el fin de encontrar el número de clúster en el cual los la media de los clúster están más separados. Se determinó que la mayor dispersión se daba con 4 clústeres. En la figura 22 se puede observar el pico de desviación estándar de las medias de SumC.

Number of Clusters	Mean Estratum	Mean Count	Mean SumC
3	0.49	1127.39	50.0299
4	0.44	858.81	50.6050
5	0.39	938.66	49.8859
6	0.36	825.79	49.4438
7	0.35	783.96	48.7309

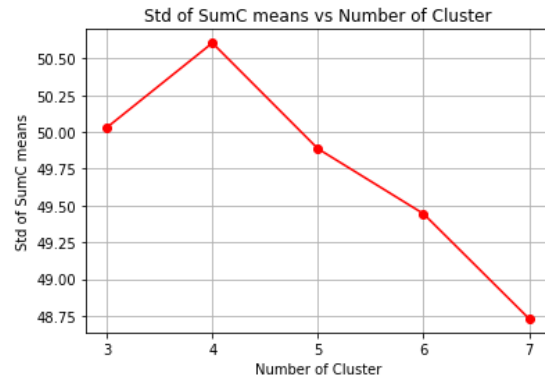


Figura 22. A la izquierda tabla de desviaciones estándar de las medias para cada cantidad de clústeres. A la derecha grafica de las desviaciones estándar de la media de SumC para cada número de clústeres.

En la figura 23 se puede apreciar la tabla con los promedios de las columnas de apreciación, la suma de las columnas de apreciación (SumC) y el estrato de para cada clúster. Se puede ver que el clúster cero corresponde a los lugares con valores más bajos en apreciación, y los valores van ascendiendo hasta llegar al clúster 4.

Cluster4	Rat	CFot	COpi	CRat	SumC	Estrato	Count
0	0.000000	3.264957	0.862714	0.000000	4.127671	3.262821	1872
1	6.992837	5.380679	2.699959	13.147360	21.227998	3.972984	2443
2	7.215855	18.505253	8.033429	32.208214	58.746896	4.125119	1047
3	7.351378	36.444882	16.358268	65.574803	118.377953	4.236220	508

Figura 23. Tabla con los promedios de las columnas de opinión, de la columna de sima de las columnas de opinión, del estrato y den conteo de cada clúster cundo se hicieron 4 clústeres.

Con lo anterior ya podríamos definir una clasificación para los lugares en Bogotá a partir de la apreciación que han dado los usuarios. En la tabla de la figura 24 podemos encontrar dicha clasificación. Se asignaron las categorías de la más baja hasta la más alta e orden alfabético iniciando por la categoría A que corresponde a la más baja hasta la categoría F siendo esta la más alta.

Clasification	Description	Count	Average Stratum	Std Straum	Kurtosis
A	All Zeros Venues	750.0	3.173333	1.197163	0.469789
B	Cluster 0 Label	1872.0	3.262821	1.102616	0.543549
C	Cluster 1 Label	2443.0	3.972984	1.101004	-0.670343
D	Cluster 2 Label	1047.0	4.125119	1.140515	-0.907376
E	Cluster 3 Label	508.0	4.236220	1.179755	-1.001810
F	Highly Appreciated Places	868.0	4.354839	1.181709	-0.962634

Figura 24. Clasificación de los lugares de Bogotá en 6 categorías

Se puede ver en la figura 25 que hay una relación directa entre el estrato promedio y la clasificación que recibieron los lugares, a medida que los lugares están clasificados en una categoría más alta el estrato es más alto, esta relación no se puede tomar estrictamente puesto que se está tomando el estrato promedio de los lugares de cada categoría y la desviación estándar comparada con la cantidad de estratos nos dice que los datos no están distribuidos en un intervalo pequeño, por el contrario nos dice que el 68.2% de los lugares de cada categoría están un intervalo de aproximadamente más y menos alrededor del estrato promedio.

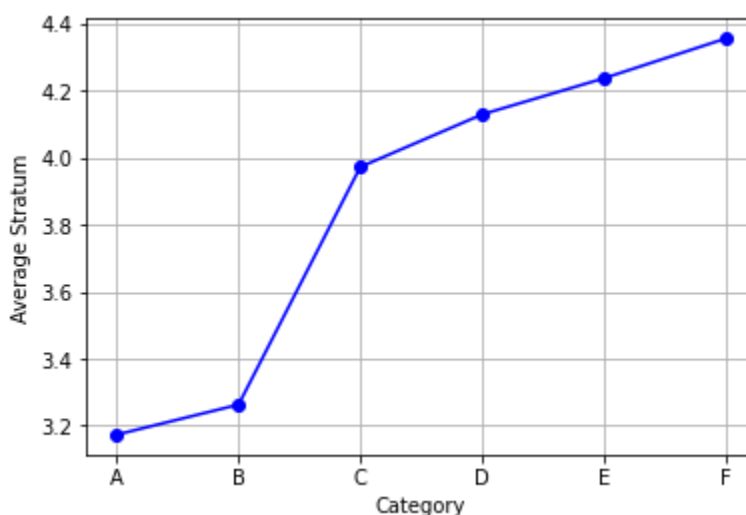


Figura 25. Estrato promedio vs categoría

4. Discusión

Se debe tener en cuenta que la base de datos que maneja foursquare no está completamente ajustada a la realidad en cuanto a número de lugares que tienen si registró, muchos lugares de marcas internacionales y en general que son de mi conocimiento no aparecen en la base de datos de foursquare.

Se ha de mencionar que se presentaron dificultades ante la falta de comprensión bajo que parámetros foursquare determina que lugares muestra a las solicitudes de lugares cercanos a un punto específico (latitud, longitud), ejemplo de esto es que buscando directamente en el la city-guía de foursquare el banco Bancolombia en Bogotá, salen tres registros mientras que en el barrido por posiciones (latitud, longitud) de la grilla triangular que cubría toda la ciudad solo apareció un registro, caso que muy seguramente ocurre con muchos otros lugares y marcas.

Por las dos razones anteriores se ha de tener en cuenta que los resultados y conclusiones encontrados no pueden ser considerados completamente certeros, pero si podemos afirmar que hay una tendencia hacia ellos.

4.1 Problema 1. Marcas Internacionales en Bogotá

Se tenía el imaginario que las marcas internacionales siempre tenderían a ubicarse en los estratos más altos con el fin de encontrar clientes con una capacidad de pago lo más alta posible, con las tendencias encontradas se puede ver que esto no es así, si hay empresas que buscan hacerse en los

más altos pero se observa que esta tendencia ocurre cuando tienen solo un punto de venta y depende también del producto y modelo de negocio de la marca.

Un fenómeno importante que no se planteó en el marco de este proyecto son los centros comerciales, debido a las mejoras en infraestructura en la ciudad se han construido centros comerciales en barrios de clase media y baja, esto permite a muchas marcas y a los usuarios entrar en contacto sin necesidad de desplazarse grandes distancias, esto se evidencia en que la mayoría de marcas encontradas en estratos 2 y 3 corresponden comidas rápidas tales como Subway, McDonald's y Dunkin' Donuts están ubicadas en centros comerciales.

Se encontró una dificultad al grande al no encontrar una base de datos abierta nacional que contenga un listado de las marcas internacionales que funcionan en el país por lo que fue necesario recurrir a listados de marcas internacionales de las cuales un porcentaje bajo fue encontrada en Bogotá, puesto que contienen muchas empresas que no funcionan en Colombia.

Un estudio más reducido a una categoría de marca permitiría identificar beneficios en la elección de la ubicación de un negocio.

4.2 Problema 2. Ubicación de los lugares en Bogotá según la opinión de los usuarios en contraste con su estrato

Se esperaba encontrar una tendencia más marcada a que los lugares con mejores apreciaciones de parte de los usuarios estuvieran ubicados en los estratos más altos, si se observó esta tendencia pero no fue tan marcada, esa tendencia esperada se pudo ver disminuida debido a que no se pensó en lugares de acceso público que suplen las necesidades de la población de forma gratuita y de alguna forma subsidiada por la ciudad, ejemplo de esto son los parques, plazas, teatros, estadios, centros comerciales y el aeropuerto, entre muchos otros que a pesar de tener una cantidad altísima de fotos, opiniones y calificaciones están ubicados en lugares que no dependen del estrato, visto de esta forma podemos considerar que la ciudad ha hecho una labor eficaz en distribuir lugares de interés público en todos los estratos.

A partir de este análisis sobre los lugares en Bogotá y la apreciación que dan sus usuarios se pueden trazar planes de mejoramiento en la atención a las necesidades de los usuarios y en los servicios que prestan los lugares según la ubicación en que se encuentra el lugar, esto puede mejorar la percepción que tienen los usuarios. Un análisis más reducido a una categoría en particular de los lugares de Bogotá junto con la apreciación que tienen los usuarios sobre este permitiría identificar en que ubicaciones es más conveniente emprender un negocio o una obra pública.

5. Conclusiones

En este informe se presentaron dos problemas relacionados con la ubicación de lugares en Bogotá y el estrato en el que se encuentran ubicados esos lugares.

El primer cuestionamiento nos llevó a encontrar una tendencia en las decisiones que toman las marcas internacionales al momento de ubicar sus puntos de venta en la ciudad, cuando la marca plantea tener solo un punto prefieren en el 60% de los casos ubicarlo en un estrato medio alto (4.1 en promedio), las marcas con entre 4 y doce puntos se ubicaron principalmente en estrato medio

(3.2 en promedio), las marcas con más de 12 puntos se ubicaron en estrato medio alto (3.9 en promedio), también es de notar que ninguna marca internacional tiene puntos en el estrato mas bajo de la ciudad. Las

En el segundo cuestionamiento se buscaba ver si existe una relación entre el estrato en que están ubicados los lugares y la apreciación que tienen los usuarios. Se clasificaron los lugares en seis categorías, siendo la más baja la A, la categoría más alta F y las categorías intermedias B, C, D y E que fueron determinadas mediante el algoritmo de clustering según su cantidad de fotos, opiniones, numero de calificaciones y calificación promedio del lugar, se encontró que el número de clústeres que separaba mejor estas cantidades era de cuatro clústeres. A partir de esta clasificación fue se obtuvo el estrato promedio de los lugares de cada categoría resultando que la categoría más baja (A) tiene el estrato más bajo en promedio (3.17 estrato medio), y a medida que aumenta la categoría aumenta el estrato, la categoría más alta (F) tiene el estrato promedio más alto (4.35 estrato medio alto). Se esperaba ver una tendencia más marcada de forma tal que en la categoría más baja el estrato fuera aproximadamente de 2 y en la categoría más alta el estrato fuera cercano a 6.

No se encontraron las tendencias tan marcadas como se esperaba debido a que hay lugares públicos como teatros, parques, plazas, centros comerciales y el aeropuerto, entre otros que prestan una experiencia agradable y de alta recordación al público independiente del estrato en el que se encuentra ubicado el lugar.

Se puede decir que si hay una tendencia a que los lugares con mejor apreciación por parte de los usuarios se encuentran en los estratos más altos pero no es la regla.