

DIRECCIÓN GEOESTADISTICA

SISTEMA DE INFORMACIÓN GEOESTADISTICA (SIGE)

**DOCUMENTO TÉCNICO DE ELABORACIÓN DE INDICADORES DEL
MGN(MARCO GEOESTADISTICO NACIONAL) A NIVEL INTEGRADO
MANZANA - CNPV2018 EN LA CIUDAD DE BOGOTÁ.**

DICIEMBRE de 2021

CONTENIDO

Procedimientos técnicos de la elaboración de indicadores a fines del déficit habitacional de acuerdo al MGN – Nivel Manzana 2018, Ciudad Bogotá DC. 11

1.	Levantamiento de datos del MGN – Nivel Manzana 2018, Ciudad Bogotá D.C.....	11
2.	Construcción de indicadores para la medición del déficit habitacional hacia la caracterización urbano-habitacional con variables del CNPV-2018.....	15
2.1	<i>Construcción del indicador: "Déficit diferencial de población con cobertura a servicios públicos".</i>	16
2.2	<i>Construcción del indicador: "Déficit en cobertura de servicios públicos" (DCSP).</i>	62
2.3	<i>Construcción del indicador: "Densidad Habitacional". (DH)</i>	86
2.4	<i>Construcción del indicador: "Garantía de desarrollo productivo"</i>	133
2.5	<i>Indicador de garantía de desarrollo productivo por sector económico:</i>	150

Trabajos citados 189

Lista de tablas

Tabla 1. Variables categóricas según el MGN-NIM2018.....	11
Tabla 2. Variables a seleccionar de tipo numérica únicas del MGN-NIM2018.	13
Tabla 3. Variables agrupadas para cada tipo de variables observadas en el MGN-NIM2018... ..	14
Tabla 4. Variables a fines a la construcción de los indicadores asociados a la cobertura a servicios públicos.....	16
Tabla 5. Resumen estadístico de variables afines para la construcción del indicador en estudio (DDPCSP).	18
Tabla 6. Resumen estadístico de variables afines para la construcción del indicador en estudio (DDPCSP).	20
Tabla 7. Resumen estadístico de aproximaciones afines para la construcción del indicador en estudio (DDPCSP).	23
Tabla 8. Resultado de aplicación de test estadísticos de Normalidad Bilateral.	27
Tabla 9. Consolidado de estadísticos aplicados a cada aproximación absoluta para el indicador DDPCSP.....	32
Tabla 10.Tabla de contingencia del indicador para cada localidad.....	35
Tabla 11. Resultados de la creación de matriz de valores de pesos para el data set MGN-NIM 2018- Bogotá DC.	42
Tabla 12. Mapas de significancia del Indicador DDPCSP con grado de verosimilitud del 99%.50	50
Tabla 13. Leyendas de Mapas de significancia del DDPCSP.	51
Tabla 14. Resumen del método de creación matriz de pesos de contiguidad.	56
Tabla 15. Gráfico de series para las manzanas en cada localidad : Niveles del DDPCSP en términos de categorías.....	60
Tabla 16. Resumen estadístico de variables afines al DCSP	62
Tabla 17. Tabla de contingencia del indicador para cada localidad.	67
Tabla 18. Mapas de clústeres del Indicador DCSP con grado de verosimilitud del 99%.....	75
Tabla 19. Leyendas de Mapas de significancia del DCSP.	75
Tabla 20. Gráfico de series para las manzanas en cada localidad : Niveles del DCSP en términos de categorías.....	83
Tabla 21. Variables a fines a la construcción de los indicadores asociado a la densidad habitacional.	86
Tabla 22. Resumen estadístico de las variables afines al indicador DH.....	88
Tabla 23. Pruebas de Hipótesis de la medición de correlación global.....	94
Tabla 24. Resumen de cantidad de factores por cada procedimiento.....	100
Tabla 25. Comparación de métodos AFE aplicados al data set.	100
Tabla 26. Análisis de intervalos de confiabilidad para cada método AFE.	102
Tabla 27.Análisis de intervalos de confiabilidad de media residuo para cada método AFE. ...	105

Tabla 28. Resultado de AFE método 1: Cargas factoriales.....	108
Tabla 29. Resultado del Alpha de Cronbach aproximaciones normales.....	115
Tabla 30. Resultado del Alpha de Cronbach aproximaciones rescaladas.	116
Tabla 31. Resumen de la creación de matriz de peso de contigüidad con el método KNN....	124
Tabla 32. Mapas de significancia : métodos LISA & Outliers en comparativa con Paralelogramo.....	132
Tabla 33. Variables afines al indicador Garantía de desarrollo productivo.	133
Tabla 34. Resumen de categorías indicador Garantía de desarrollo 1.	137
Tabla 35. Resumen de la creación de matriz de peso de contigüidad con el método KNN....	141
Tabla 36. Valores de sensibilidad del alfa de cronbach (raw_alfa)	154
Tabla 37. Resumen de categorías de las variables a fines al GD2.....	156
Tabla 38. Comparación de categorías en función de los códigos de manzanas por localidad de los sectores económicos calculados.....	157
Tabla 39. Criterios de clasificación de la prueba de adecuación muestral KMO.....	165
Tabla 40. Métodos disponible (Total :19) para el Análisis Factorial Exploratorio.	167
Tabla 41. Resultado de AFE : Cargas factoriales.	169
Tabla 42. Valores de las normalizaciones para cada sector económico (ponderación o peso)	174
Tabla 43. Resumen de categorías indicador Garantía de desarrollo 2.....	176

Lista de imágenes

Imagen 1. Captura de la función de resumen de datos para el data set MGN-NIM 2018	11
Imagen 2.Sentencia para filtrar la ciudad Bogotá DC en MGN-NIM2018	12
Imagen 3.Captura de la función de resumen de datos para el data set filtrado Bogotá D.C en MGN-NIM 2018.....	12
Imagen 4. Captura de la función de estructura de datos para el data set filtrado Bogotá D.C en MGN-NIM 2018.....	14
Imagen 5. Captura de la función de resumen de datos para las variables seleccionadas y el data set filtrado Bogotá D.C en MGN-NIM 2018.....	14
Imagen 6. Resultado de fracción de la primera aproximación (SPF vs SPNF) del indicador DDPCSP.....	24
Imagen 7. Resultado de fracción de la segunda aproximación (PPF vs PPNF) del indicador DDPCSP.....	25

Lista de gráficos

Gráfico 1. Matriz de correlaciones variables a fin del DDPCSP	21
Gráfico 2. Gráficos Q-Q para cada aproximación del indicador DDPCP.....	26
Gráfico 3. Histograma con curva normal teórica de la primera aproximación absoluta del DDPCSP.....	29
Gráfico 4. Histograma con curva normal teórica de la segunda aproximación absoluta del DDPCSP.....	30
Gráfico 5. Gráfico de densidad del indicador DDPCSP por localidad en la ciudad de Bogotá DC.....	33
Gráfico 6. Scree plot: PCA gráfico de barras de componentes.....	36
Gráfico 7. Gráfico de tabla cruzadas del indicador DDPCSP.....	38
Gráfico 8. Gráfico de dispersión del índice de Morán (DDIM) : DDPCSP.	43
Gráfico 9. Palalegrama, Matriz de correlaciones afines DDPCSP.....	49
Gráfico 10. Gráficos de dispersión Índice de Morán.	58
Gráfico 11. Histograma de datos SPS.....	64
Gráfico 12. Histograma del indicador DCSP.....	65
Gráfico 13. Gráfico de densidad del indicador DCSP por localidad en la ciudad de Bogotá DC.	66
Gráfico 14. Scree plot: PCA gráfico de barras de componentes principales.	68
Gráfico 15. Gráfico de tabla cruzada del indicador DCSP.....	70
Gráfico 16. Gráfico de dispersión del índice de Morán : DCSP.	74
Gráfico 17. Histograma del indicador DCSP en la ciudad de Bogotá DC.....	79
Gráfico 18. Gráfico de dispersión DCSP vs DDPCSP	84
Gráfico 19. Gráfico de dispersión filtrado para valores mayores en ambos indicadores a 0.8..	84
Gráfico 20. Análisis de densidad de kernel: Diagrama dispersión DCSP vs DDPCSP	85
Gráfico 21.Histogramas de las aproximaciones del indicador DH.....	89
Gráfico 22. Gráficos de cajas de las aproximaciones del indicador DH	89
Gráfico 23. Gráfico de Q-Q de las dos aproximaciones.....	90
Gráfico 24. Matriz de correlaciones de las variables afines al indicador DH.....	93
Gráfico 25. Gráfico de barras de consenso de métodos en función de Factores a extraer.	97
Gráfico 26. Gráfico método Scree AFE	98
Gráfico 27. Gráfico Scree bajo el método Parallel.....	99
Gráfico 28. Gráfico de análisis Parallel.....	99
Gráfico 29. Matriz de dispersión entre factores.....	111
Gráfico 30. Gráfico de densidad del indicador DH por localidad.....	117
Gráfico 31. Gráfico de tablas cruzadas del indicador Densidad Habitacional por categorías.	119

Gráfico 32. Gráfico del Índice de Morán Indicador DH.....	125
Gráfico 33. Gráfico de densidad del indicador DH.....	129
Gráfico 34. Gráfico de densidad del indicador GD1 por localidad	136
Gráfico 35. Gráfico de tablas cruzadas del indicador Garantía de desarrollo	140
Gráfico 36. Gráfico del Índice de Morán Indicador DH.....	142
Gráfico 37. Gráfico de densidad del indicador GD1.....	147
Gráfico 38. Histograma de cada proporción de sector económico asociado y la variable PROD.	155
Gráfico 39. Gráfico de tablas cruzadas de los sectores económicos.....	161
Gráfico 40. Matriz de correlaciones de las variables afines al indicador GD2.	163
Gráfico 41. Pruebas de Hipótesis de la medición de correlación global variables afines GD2.	164
Gráfico 42. Gráfico de barras de consenso de métodos en función de Factores a extraer.	166
Gráfico 43. Gráfico método Scree AFE	168
Gráfico 44. Gráfico de análisis Parallel.....	168
Gráfico 45. Gráfico de densidad del indicador GD2 por localidad	175
Gráfico 46. Gráfico de tablas cruzadas del indicador Garantía de desarrollo 2.	179

Lista de diagramas

Diagrama 1. PCA método circular de tabla cruzada.	37
Diagrama 2. PCA método gráfico de tablas cruzada	37
Diagrama 3. Diagrama de árbol de participación por categoría y por localidad	39
Diagrama 4. PCA método circular de tabla cruzada.	68
Diagrama 5. PCA método gráfico de tablas cruzada	68
Diagrama 6. Diagrama Biplot Análisis de componentes principales.....	69
Diagrama 7. Diagrama de árbol de participación por categoría y por localidad DCSP.....	71
Diagrama 8. Análisis factorial exploratorio gráfico.....	110
Diagrama 9. Diagrama ICLUST indicador Densidad Habitacional.....	112
Diagrama 10. Modelo bifactorial Omega.....	113
Diagrama 11.PCA Método Gráfico Espacial : índice DH	118
Diagrama 12. PCA método circular.	119
Diagrama 13. Diagramas de tablas cruzadas del indicador Densidad Habitacional por categorías.	120
Diagrama 14.PCA Método Gráfico Espacial : GD1	138

Diagrama 15. PCA método circular.....	139
Diagrama 16. Diagrama de árbol de las tablas cruzadas en el indicador Garantía de desarrollo.....	140
Diagrama 17. Diagrama de árbol de las tablas cruzadas para los sectores económicos.....	162
Diagrama 18. Análisis factorial exploratorio gráfico.....	171
Diagrama 19. PCA Método Gráfico Espacial individual por localidades : GD2.....	177
Diagrama 20. PCA método circular.....	178
Diagrama 21. Diagrama de árbol de las tablas cruzadas en el indicador Garantía de desarrollo 2.....	180

Lista de ilustraciones.

Ilustración 1. Resultado operativo de sumatoria por columna de cada aproximación a fin para la construcción del indicador en estudio (DDPCSP).....	24
Ilustración 2. Resultado de aplicación de test estadístico Fligner-Killeen para evaluación de Homocesdasticidad entre representaciones de poblaciones (PF vs PNF).	28
Ilustración 3. Resumen de aplicación test Levene Homogeneidad de varianza del indicador por cada localidad.....	31
Ilustración 4. Mapa de calor de la tabla de contingencia del indicador DDPCSP	35
Ilustración 5. Mapa distribución geoespacial de las manzanas (datos vectoriales) en la ciudad de Bogotá DC por localidad registrada (Gráfico geoespacial único de localidades).....	40
Ilustración 6. Mapa de distribución geoespacial del indicador DDPCSP en Cuartiles. (5).....	41
Ilustración 7. Mapa de distribución geoespacial del indicador DDPCSP en Cuartiles. (1.5 Rango Intercuartílico). (Gráfico de clasificación intercuartílico).....	42
Ilustración 8. Mapas de contraste del DDIM con el gráfico único de localidades.....	44
Ilustración 9. Mapas de contraste del DDIM con el gráfico de clasificación intercuartilica.....	45
Ilustración 10. Mapa de significancia y Clúster : Método de “Índice de Morán y LISA”.....	46
Ilustración 11. Mapa de significancia y Clúster : Método de “estadístico Getis-ord Gi*”.....	47
Ilustración 12. Mapa de significancia y clúster : Método “Local Geary Clúster”	48
Ilustración 13. Representación de resultado de auto correlación espacial.....	51
Ilustración 14. Mapas de significancia estadística método Gi* vs Mapa único de localidades.	52
Ilustración 15. Análisis de las manzanas con mediciones elevadas del indicador : DDPCSP	53
Ilustración 16. Mapa localidad Teusaquillo con distribución geoespacial del indicador DDPCSP.	54
Ilustración 17. Mapas de selección de manzanas en el barrio “La Esmeralda”	55
Ilustración 18. Mapa de interpolación kriging del indicador DDPCSP.	55

Ilustración 19. Mapa de significancia estadística método LISA : Localidad Teusaquillo.....	56
Ilustración 20. Mapa de clúster método LISA : Localidad Teusaquillo.	57
Ilustración 21. Mapas de desviación estándar del índice DDPCSP en el Barrio : "La Esmeralda".	59
Ilustración 22. Mapa de clúster localidad teusaquillo (Operación promedios entre manzanas).	61
Ilustración 23. Mapas de clústeres de 5 clústeres.	61
Ilustración 24. Resumen de cantidad de manzanas ubicadas en cada categoría DCSP	66
Ilustración 25. Mapa de calor de la tabla de contingencia del indicador DCSP.....	67
Ilustración 26. Mapa de distribución geoespacial del indicador DCSP en Cuartiles. (5)	72
Ilustración 27. Mapa de distribución geoespacial del indicador DCSP en Cuartiles. (1.5 Rango Intercuartílico). (Gráfico de clasificación intercuartílico).....	73
Ilustración 28. Mapas de clústeres del método LISA vs Mapa único de localidades vs Mapas único de Categorías del Indicador DCSP.	76
Ilustración 29. Mapas condicionales bivariado: DDPCSP vs DCSP	77
Ilustración 30. Análisis de las manzanas con mediciones elevadas del indicador : DCSP	78
Ilustración 31. Mapa localidad Teusaquillo con distribución geoespacial del indicador DCSP.	79
Ilustración 32. Mapas de selección de manzanas en el barrio "La Esmeralda"	80
Ilustración 33. Mapa de interpolación kriging del indicador DCSP.....	80
Ilustración 34. Mapa de significancia estadística método LISA : Localidad Teusaquillo.	81
Ilustración 35. Mapa de clúster método LISA : Localidad Teusaquillo.	82
Ilustración 36. Resultado de la prueba Fligner-Killeen del Homogeneidad de Varianzas.	91
Ilustración 37. Resultado de la prueba Levene del Homogeneidad de Varianzas para las localidades.	91
Ilustración 38. Criterios de clasificación de la prueba de adecuación muestral KMO.	96
Ilustración 39. Reporte del consenso entre métodos AFE.	97
Ilustración 40. Métodos disponible (Total :14) para el Análisis Factorial Exploratorio.	97
Ilustración 41. Resultados del test BIC y Chi-Squared para cada modelo.	107
Ilustración 42. Reporte de resultados del AFE: Método 1.....	109
Ilustración 43. Calculo de las cargas factoriales por cada variable (Agrupación en términos de peso).....	114
Ilustración 44. Análisis de Alpha de Cronbach: Intervalo de confianza.....	114
Ilustración 45. Resumen de categorías indicador Densidad Habitacional	117
Ilustración 46. Mapa de calor del indicador Densidad Habitacional.....	118
Ilustración 47. Mapa de distribución geoespacial del indicador DH en Cuartiles (5)	121
Ilustración 48. Mapas de distribución geoespacial del indicador DH en Cuartiles (1.5 veces el Rango Intercuartilico). (Gráfico de clasificación intercuartílico)	122

Ilustración 49. Mapas segmentados de comparación de poblaciones con valores en Alto Superávit de Densidad Habitacional.....	122
Ilustración 50. Mapas comparativos de poblaciones con alta desviación estándar indicador DH.....	123
Ilustración 51. Mapas de significancia y Clúster : Método de “Índice de Morán y LISA” 95% confiabilidad y 999 permutaciones.....	125
Ilustración 52. Mapas comparativos de significancia estadística con un 99% de confiabilidad ColdSpots	127
Ilustración 53. Análisis de las manzanas con mediciones elevadas del indicador : DH.....	128
Ilustración 54. Mapa localidad Teusaquillo con distribución geoespacial del indicador DCSP.....	129
Ilustración 55. Mapa de significancia método LISA, 95% confiabilidad para el indicador DH.	130
Ilustración 56. Mapa de clúster por el método LISA con peso de contigüidad KNN.	131
Ilustración 57. Resumen estadístico de las variables agrupadas por categoría productiva.	135
Ilustración 58. Resumes de variables empleadas para el indicador garantía de desarrollo 1.	135
Ilustración 59. Mapa de calor del indicador Garantía desarrollo 1.....	137
Ilustración 60. Mapas de significancia y Clúster : Método de “Índice de Morán y LISA” 95% confiabilidad y 999 permutaciones.....	143
Ilustración 61. Mapas comparativos de significancia estadística con un 99% de confiabilidad ColdSpots	145
Ilustración 62. Análisis de las manzanas con mediciones elevadas del indicador : GD1.	146
Ilustración 63. Mapa localidad Teusaquillo con distribución geoespacial del indicador GD1.	147
Ilustración 64. Mapa de significancia método LISA, 95% confiabilidad para el indicador GD1.....	148
Ilustración 65. Mapa de clúster por el método LISA con peso de contigüidad KNN del indicador GD1.....	149
Ilustración 66. Resumen estadístico de variables por tipo de actividad afines al indicador de “Garantía de desarrollo por sector económico”.....	151
Ilustración 67. Resumen estadístico integrando las tres variables de sectores económicos. ..	151
Ilustración 68. Matriz de correlaciones variables afines al indicador: “Garantía de desarrollo 2” (GD2).	152
Ilustración 69. Correlograma de matriz de correlaciones de variables afines al indicador: GD2.....	153
Ilustración 70. Resumen del alfa de cronbach variables afines al GD2.	153
Ilustración 71. Resumen de las proporciones por cada sector económico.	154
Ilustración 72. Mapa de calor de los sectores económicos.	159
Ilustración 73. Reporte del consenso entre métodos AFE.	166

Ilustración 74. Reporte de resultados del AFE.....	170
Ilustración 75.Calculo de las cargas factoriales por cada variable (Agrupación en términos de peso).....	172
Ilustración 76. Pesos factoriales pre-seleccionados para la construcción del GD2.	173
Ilustración 77. Resumen estadístico con estimación del indicador "Garantía de desarrollo por sector económico".	174
Ilustración 78. Mapa de calor del indicador Garantía desarrollo 2.	176
Ilustración 79. Mapa de distribución geoespacial de las categorías del índice GD2	181
Ilustración 80. Mapas de significancia y Clúster : Método de "Índice de Morán y LISA" 95% confiabilidad y 999 permutaciones.....	182
Ilustración 81. Mapas comparativos de significancia estadística con un 95% de confiabilidad HotSpots	183
Ilustración 82. Mapas comparativos de significancia estadística con un 95% de confiabilidad ColdSpots	184
Ilustración 83.Análisis de las manzanas con mediciones elevadas del indicador : GD2.	185
Ilustración 84. Mapa localidad Teusaquillo con distribución geoespacial del indicador GD2.186	
Ilustración 85. Mapa de significancia método LISA, 95% confiabilidad para el indicador GD2.	187
Ilustración 86. Mapa de clúster por el método LISA con peso de contigüidad KNN para el indicador GD2.....	188

Procedimientos técnicos de la elaboración de indicadores a fines del déficit habitacional de acuerdo al MGN – Nivel Manzana 2018, Ciudad Bogotá DC.

1. Levantamiento de datos del MGN – Nivel Manzana 2018, Ciudad Bogotá D.C.

El set de datos relacionado con el Marco Geoestadístico Nacional es el resultado del Censo Nacional de Población y vivienda -CNPV 2018 encontrado en el geo portal del DANE, el cual permite deslumbrar a ciencia cierta la información integrada de las principales variables estadísticas de la población colombiana. En mayor detalle, es posible señalar que el censo de población y vivienda consistió tanto en caracterizar las personas residentes en Colombia, como las viviendas y los hogares permitiendo resolver interrogantes importantes para la medición de indicadores demográficos, tales como: "¿Cuántas personas habitan en Colombia?, ¿Dónde están ubicadas?, ¿Cómo viven estas personas?", así mismo, esta información del censo permite tomar decisiones significativas en cualquier sector de la economía.

El volumen de variables que asciende este data-set es de 110 en total, las cantidades de manzanas (observaciones) ascienden a un total de 504.996 para todo el país, como es posible observar en el siguiente resultado una vez procesado el data-set.

Imagen 1. Captura de la función de resumen de datos para el data set MGN-NIM 2018

La cantidad de datos alojados en el archivos MGN – Manzanas 2018 es de : (504996, 110)

Fuente: Elaboración propia. Observatorio Inmobiliario Nacional DANE

Considerando el diccionario que proporciona información sobre el data-set en procesamiento, a continuación, se muestran las concepciones y significados de cada variable que explican los atributos de la población, respectivamente:

+ Las variables de tipo carácter son:

Tabla 1. Variables categóricas según el MGN-NIM2018

VARIABLE	TIPO	LONGITUD	DESCRIPCIÓN	Categoría original
COD_DANE_A	Text	22	Código de manzana concatenado (departamento, municipio, clase, sector rural, sección rural, zona urbana, sector urbano, sección urbana y manzana)	
DPTO_CCDGO	Text	2	Código del departamento	
MPIO_CCDGO	Text	3	Código del municipio	
MPIO_CDPMP	Text	5	Código concatenado que identifica al municipio	

CLAS_CCDGO	Text	1	Código de la clase 1 cabecera municipal, 2 centro poblado y 3 área resto municipal	
SETR_CCDGO	Text	3	Código de sector rural	
SETR_CCNCT	Text	9	Código de sector rural concatenado (departamento, municipio, clase y sector rural)	
SECR_CCDGO	Text	2	Código de sección rural	
SECR_CCNCT	Text	11	Código de sección rural concatenado (departamento, municipio, clase, sector rural y sección rural)	
ZU_CCDGO	Text	3	Código zona urbana (cabecera municipal y centro poblado)	
ZU_CDIVI	Text	8	Código zona urbana concatenado (departamento, municipio y zona urbana)	
SETU_CCDGO	Text	4	Código sector urbano	
SETU_CCNCT	Text	18	Código sector concatenado (departamento, municipio, clase, sector rural, sección rural, zona urbana y sector rural)	
SECU_CCDGO	Text	2	Código sección urbana	
SECU_CCNCT	Text	20	Código sección urbana concatenado (departamento, municipio, clase, sector rural, sección rural, zona urbana, sector urbano y sección urbana)	
MANZ_CCDGO	Text	2	Código de manzana	
AG_CCDGO	Text	6	Código Área Geográfica AG	
DATO_ANM	Text	50	Nombre capa anonimizada	
CD_LC_CM	Text	10	Código de la localidad o comuna	
NMB_LC_CM	Text	50	Nombre de la localidad o comuna	
TP_LC_CM	Text	20	Descripción de tipo localidad o comuna o corregimiento	

Fuente: Diccionario MGN-NIM2018. Observatorio Inmobiliario Nacional DANE

El estudio en este documento se concentrará en la ciudad de Bogotá D.C, esta tiene el código número 11 según la nomenclatura para las ciudades, por lo que se procede a filtrar este dataset para la ciudad a través de Python, enunciando la siguiente sentencia:

Imagen 2.Sentencia para filtrar la ciudad Bogotá DC en MGN-NIM2018

```
df_1 = df[df['DPTO_CCDGO'] == '11']
```

Fuente: Elaboración propia a través de python. Observatorio Inmobiliario Nacional DANE

Se evidencia que para la ciudad de Bogotá DC la cantidad de registros son 43408 datos.

Imagen 3.Captura de la función de resumen de datos para el data set filtrado Bogotá D.C en MGN-NIM 2018

```
La cantidad de datos alojados en el archivos MGN – Manzanas 2018 es de : (43408, 110)
```

Fuente: Elaboración propia a través de python. Observatorio Inmobiliario Nacional DANE

Las variables refieren interés para la construcción de indicadores son las que proporcionan información única y escalable, es decir que permitan ser utilizadas para filtro emitiendo así hallazgos en términos de detalles para la población y respectiva caracterización, razón por la cual se emplea la selección de las siguientes variables:

Tabla 2. Variables a seleccionar de tipo numérica únicas del MGN-NIM2018.

VARIABLE	TIPO	LONGITUD	DESCRIPCIÓN
COD_DANE_A	Text	22	Código de manzana concatenado (departamento, municipio, clase, sector rural, sección rural, zona urbana, sector urbano, sección urbana y manzana)
DPTO_CCDGO	Text	2	Código del departamento
MPIO_CCDGO	Text	3	Código del municipio
MPIO_CDPMP	Text	5	Código concatenado que identifica al municipio
CLAS_CCDGO	Text	1	Código de la clase 1 cabecera municipal, 2 centro poblado y 3 área resto municipal
SETR_CCDGO	Text	3	Código de sector rural
SETR_CCNCT	Text	9	Código de sector rural concatenado (departamento, municipio, clase y sector rural)
SECR_CCDGO	Text	2	Código de sección rural
SECR_CCNCT	Text	11	Código de sección rural concatenado (departamento, municipio, clase, sector rural y sección rural)
ZU_CCDGO	Text	3	Código zona urbana (cabecera municipal y centro poblado)
ZU_CDIVI	Text	8	Código zona urbana concatenado (departamento, municipio y zona urbana)
SETU_CCDGO	Text	4	Código sector urbano
SETU_CCNCT	Text	18	Código sector concatenado (departamento, municipio, clase, sector rural, sección rural, zona urbana y sector rural)
SECU_CCDGO	Text	2	Código sección urbana
SECU_CCNCT	Text	20	Código sección urbana concatenado (departamento, municipio, clase, sector rural, sección rural, zona urbana, sector urbano y sección urbana)
MANZ_CCDGO	Text	2	Código de manzana
AG_CCDGO	Text	6	Código Área Geográfica AG
DATO_ANM	Text	50	Nombre capa anonimizada
CD_LC_CM	Text	10	Código de la localidad o comuna
NMB_LC_CM	Text	50	Nombre de la localidad o comuna
TP_LC_CM	Text	20	Descripción de tipo localidad o comuna o corregimiento

Fuente: Diccionario MGN-NIM2018. Observatorio Inmobiliario Nacional DANE

Los colores indican: Las variables de color azul , corresponden a las variables que cumplen con la condición de variables únicas, permitiendo evitar duplicidad y mayor apropiación de la información en términos de categoría de población.

Empleando la selección, se obtiene que en total en el data-set, contiene 87 variables numéricas de tipo continua, 11 variables numéricas de tipo entero y 5 de tipo categórica, como se muestra a continuación:

Imagen 4. Captura de la función de estructura de datos para el data set filtrado Bogotá D.C en MGN-NIM 2018

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43408 entries, 0 to 43407
Columns: 103 entries, DPTO_CCDGO to geometry
dtypes: float64(87), int64(11), object(5)
memory usage: 34.1+ MB
```

Fuente: Elaboración propia a través de python. Observatorio Inmobiliario Nacional DANE

La selección da como resultado que 103 variables en total, las cuales serían útiles para la construcción de los indicadores que permitan caracterizar y explicar las condiciones habitacionales de la población en la ciudad de Bogotá D.C.

Imagen 5. Captura de la función de resumen de datos para las variables seleccionadas y el data set filtrado Bogotá D.C en MGN-NIM 2018

```
La cantidad de datos alojados en el archivos MGN – Manzanas 2018
despues de seleccionar las variables unicas tipo texto es de : (43408, 103)
```

Fuente: Elaboración propia a través de python. Observatorio Inmobiliario Nacional DANE

+ Las variables que tiene valores numéricos agrupadas en términos de proporciones se muestran a continuación:

Tabla 3. Variables agrupadas para cada tipo de variables observadas en el MGN-NIM2018.

Grupo de variable	%
Conteo de viviendas con tipo de servicio	24,7%
Conteo de personas	22,4%
Conteo de unidades no residenciales	12,9%
Conteo de viviendas tipo	12,9%
Cantidad de encuestas	8,2%
Conteo de unidades mixtas con uso	5,9%
Conteo de unidades con uso	4,7%
Año información geográfica	1,2%
Área de manzana en metros cuadrados	1,2%

Conteo de hogares	1,2%
Latitude	1,2%
Longitude	1,2%
Número de personas	1,2%
Número promedio de habitantes en la manzana por metro cuadrado	1,2%

Fuente: Diccionario MGN-NIM2018, Elaboración desde EXCEL, Observatorio Inmobiliario Nacional DANE

El 91% de las variables que contiene el data-set sobre 87 variables que presentan datos agrupados en una misma medición, explican los datos en información de caracterización alusivas a ser asociadas o articuladas a indicadores que permitan llegar a conclusiones sobre la población de forma relacional con las distribuciones geográficas, esto quiere decir que gracias a que existen más proporción de variables de tipo numéricas es posible caracterizar en términos cuantitativos la población.

2. Construcción de indicadores para la medición del déficit habitacional hacia la caracterización urbano-habitacional con variables del CNPV-2018.

La construcción de indicadores que permitan medir el déficit habitacional visto desde diferentes ángulos, permite aproximar un análisis crítico del valor de forma que integra la explicación mediante las cifras obtenidas luego de una correcta medición, estas capaces de demostrar según la distribución del resultado del CNPV, las zonas en las que predominan cifras afines al entendimiento de la población con parámetros estacionales de acuerdo a sus condiciones de entorno, sociales, de seguridad y de diseño estructural, que permiten caracterizar en detalle cada división político administrativa y áreas geo estadísticas.

Existen dos tipos de déficit, conforme a la necesidad tanto de estimar si la cantidad de viviendas que existen es suficiente para albergar la totalidad de la población del país (déficit cuantitativo), y en segundo lugar, la necesidad de cuantificar el numero de viviendas que no cumplen con estándares mínimos de calidad de vida adecuada con los moradores (déficit cualitativo). Conforme al déficit habitacional entonces se definen los dos tipos:

Déficit cuantitativo:

Estima la cantidad de viviendas que la sociedad debe construir o adicionar para que exista una relación uno a uno entre viviendas adecuadas y hogares que necesitan ser alojados, de forma absoluta refiere a la diferencia entre el número de hogares y las viviendas.

Déficit cualitativo:

Aproximación de indicadores que regulen la satisfacción a las condiciones de vida de vivienda y la relación con el entorno que permita el desarrollo social, los atributos que dignifican a un individuo por la concatenación de beneficios y márgenes aceptables de crecimiento personal, dotados del mantenimiento de condiciones de higiene y salubridad, alusivos a las mejores condiciones mínimas de habitabilidad para los moradores.

2.1 Construcción del indicador: “Déficit diferencial de población con cobertura a servicios públicos”.

El déficit diferencial de población con cobertura a servicios públicos, refiere de la incidencia de la población con cobertura a servicios contra la población sin cobertura o conexión a redes publicas, es decir la diferencia entre las poblaciones en cada manzana registrada en el dataset MGN-NIM2018 que tienen acceso a los servicios públicos con respecto a la población en estado de carencia en la satisfacción de las necesidades de acceso a servicios. Este indicador permite medir en que zonas existen mayoría de poblaciones favorecidas o que cuentan con excelentes condiciones habitacionales y en que zonas entendiéndose de las localidades y manzanas de la ciudad de Bogotá DC, existe minoría de poblaciones favorecidas, estas mediciones contempladas gracias a las aproximaciones por límites geográficos.

+ Relacionamiento de conteo de viviendas con diferentes tipos de servicios.

Los índice en cobertura de servicios públicos permiten medir la incidencia de estados de carencia en la satisfacción completa de las necesidades de la población organizada por manzana integrada de acuerdo al acceso a servicios públicos, construido a partir de las siguientes 22 variables. Las variables corresponde a información de frecuencia de conteo de viviendas con características de favorecimiento, por contar con servicios públicos y conteo de viviendas que no cuentan con servicios públicos. Las variables se muestran a continuación:

Tabla 4. Variables a fines a la construcción de los indicadores asociados a la cobertura a servicios públicos.

VARIABLE	INDICADOR	CLASE	DESCRIPCIÓN	Categoría original
TP19_EE_1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas con energía eléctrica	Si
TP19_EE_2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin energía eléctrica	No
TP19_EE_E1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en Estrato 1	Estrato 1
TP19_EE_E2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en Estrato 2	Estrato 2

NOMBRE DE LA INVESTIGACIÓN / TEMÁTICA

TP19_EE_E3	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en Estrato 3	Estrato 3
TP19_EE_E4	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en Estrato 4	Estrato 4
TP19_EE_E5	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en Estrato 5	Estrato 5
TP19_EE_E6	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en Estrato 6	Estrato 6
TP19_EE_E9	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas que reportan recibir facturación de energía eléctrica en No sabe o no tiene estrato	No sabe o no tiene estrato
TP19_ACU_1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas con servicio de acueducto	Si
TP19_ACU_2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin servicio de acueducto	No
TP19_ALC_1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas con servicio de alcantarillado	Si
TP19_ALC_2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin servicio de alcantarillado	No
TP19_GAS_1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas con servicio de gas natural conectado a red pública	Si
TP19_GAS_2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin servicio de gas natural conectado a red pública	No
TP19_GAS_9	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin información de servicio de gas natural conectado a red pública	Sin información
TP19_REC_B1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas con servicio de recolección de basuras	Si
TP19_REC_B2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin servicio de recolección de basuras	No
TP19_INTE1	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas con servicio de internet	Si
TP19_INTE2	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin servicio de internet	No
TP19_INTE9	DEFICIT SERVICIOS PUB	Conteo de viviendas con tipo de servicio	Conteo de viviendas sin información de servicio de internet	Sin información
TVIVIENDA	DEFICIT SERVICIOS PUB	Conteo de viviendas tipo	Conteo de viviendas	

Fuente: Diccionario MGN-NIM2018, Elaborado desde EXCEL, Observatorio Inmobiliario Nacional DANE

En ese orden ideas se presentan a continuación los análisis elaborados en función de la construcción de cada indicador asociado con el diferencial de población con cobertura a servicios públicos de la población de Bogotá D.C.

La escala de los datos para la construcción de este indicador, se regula según la implementación las operaciones a fines para la agrupación y relacionamiento de las variables seleccionadas, con la finalidad de destacar con atributos las poblaciones donde existe minoría y mayoría.

Tabla 5. Resumen estadístico de variables afines para la construcción del indicador en estudio (DDPCSP).

Data summary	
Name	df_1
Number of rows	43408
Number of columns	21
<hr/>	
Column type frequency:	
numeric	21
<hr/>	
Group variables	None

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
TP19_EE_1	0	1	53.80	92.85	0	16	35	60	2844	
TP19_EE_2	0	1	0.13	1.10	0	0	0	0	102	
TP19_EE_E1	0	1	4.49	14.58	0	0	0	1	672	
TP19_EE_E2	0	1	21.11	57.79	0	0	1	28	2795	
TP19_EE_E3	0	1	18.66	58.56	0	0	0	16	1659	
TP19_EE_E4	0	1	6.07	45.30	0	0	0	0	2059	
TP19_EE_E5	0	1	1.92	19.38	0	0	0	0	1212	
TP19_EE_E6	0	1	1.41	12.72	0	0	0	0	485	
TP19_EE_E9	0	1	0.14	0.81	0	0	0	0	84	
TP19_ACU_1	0	1	53.68	92.92	0	16	34	60	2842	
TP19_ACU_2	0	1	0.26	2.33	0	0	0	0	214	
TP19_ALC_1	0	1	53.55	92.87	0	15	34	60	2844	
TP19_ALC_2	0	1	0.38	2.86	0	0	0	0	219	
TP19_GAS_1	0	1	50.29	90.16	0	14	32	56	2830	
TP19_GAS_2	0	1	2.97	9.04	0	0	1	3	420	
TP19_GAS_9	0	1	0.68	2.31	0	0	0	1	320	
TP19_REC_B1	0	1	53.63	92.75	0	16	34	60	2846	
TP19_REC_B2	0	1	0.30	2.00	0	0	0	0	186	
TP19_INTE1	0	1	40.27	80.40	0	9	23	42	2230	
TP19_INTE2	0	1	12.98	19.64	0	2	8	17	1300	
TP19_INTE9	0	1	0.68	2.32	0	0	0	1	320	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE

Conforme a la **Tabla 5**, la distribución de datos para cada variable agrupada, presentan un sesgo a la derecha lo que refiere que mucho de los datos se concentran con valores inferiores, la cantidad de registros que representan máximos ascienden a magnitudes que el conteo de viviendas es mayor a 2000. Conforme a los valores y su partición porcentual con cuartiles, se evidencia que solo las variables con registros máximos son las que presentan medidas estadísticas elevadas en general, la mayor cantidad de valores de viviendas con características de favorecimiento de acuerdo a los servicios ascienden en la concentración del 25% del total de manzanas en la ciudad de Bogotá D.C a 16.

Relacionamiento de conteo de viviendas con diferentes tipos de servicios (DDPCSP).

Selección de las variables influyentes en el indicador:

Se seleccionan las variables con información única y representativa, en este caso por tipo de servicio en función de las cantidades de viviendas, estas se distribuyen como lo muestra el histograma de la **Tabla 6**, con un sesgo a la derecha, y con máximos valores para las poblaciones favorecidas.

Tabla 6. Resumen estadístico de variables afines para la construcción del indicador en estudio (DDPCSP).

— Data Summary —		Values								
Name	df_3									
Number of rows	43408									
Number of columns	15									
Column type frequency:										
numeric	15									
Group variables		None								
— Variable type: numeric —										
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1 TP19_EE_1	0	1	53.8	92.9	0	16	35	60	2844	
2 TP19_EE_2	0	1	0.129	1.10	0	0	0	0	102	
3 TP19_EE_E9	0	1	0.138	0.814	0	0	0	0	84	
4 TP19_ACU_1	0	1	53.7	92.9	0	16	34	60	2842	
5 TP19_ACU_2	0	1	0.257	2.33	0	0	0	0	214	
6 TP19_ALC_1	0	1	53.6	92.9	0	15	34	60	2844	
7 TP19_ALC_2	0	1	0.381	2.86	0	0	0	0	219	
8 TP19_GAS_1	0	1	50.3	90.2	0	14	32	56	2830	
9 TP19_GAS_2	0	1	2.97	9.04	0	0	1	3	420	
10 TP19_GAS_9	0	1	0.676	2.31	0	0	0	1	320	
11 TP19_RECBI	0	1	53.6	92.7	0	16	34	60	2846	
12 TP19_RECBI2	0	1	0.304	2.00	0	0	0	0	186	
13 TP19_INTE1	0	1	40.3	80.4	0	9	23	42	2230	
14 TP19_INTE2	0	1	13.0	19.6	0	2	8	17	1300	
15 TP19_INTE9	0	1	0.684	2.32	0	0	0	1	320	

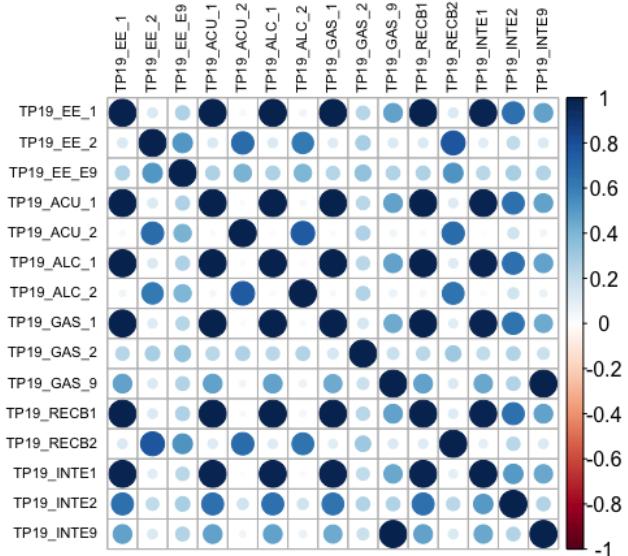
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE

Según el resumen estadístico, es posible apreciar que el máximo valor de la desviación típica es 92.9, el mínimo es 0.81 (Población que no genera información sobre alguna instancia con cobertura de servicio) estos valores estadístico permiten inferir que para las poblaciones en las que se demanda servicios básicos tienen el mismo nivel de poblaciones de acuerdo a sus valores de tendencia central a diferencia de la población que tiene internet esta corresponde a un 23,93% menos que la población regular que tiene todos los servicios. A continuación, se procede a aplicar operaciones entre variables para determinar el indicador de DDPCSP.

Ilustración de entidades de correlación variables del DDPCSP

Analizando la matriz de correlaciones, se encuentra un patrón correlacional, en el que las variables a partir del conteo de viviendas con servicios conectados y para todos los servicios, presentan alto grado de correlación positiva. Por tanto un aumento de valores para cualesquiera de este tipo de variables de viviendas favorecidas, representa un posible paralelismo en términos de dirección positiva para las variables que pertenezcan al conteo de viviendas con servicio de redes (favorecidas) conectado diferente al analizado, es decir que conforme a están distribuidos los datos analizados se presenta correlación entre el conteo de viviendas con servicio a modo de ejemplo, se hace referencia al las que cuentan con servicio eléctrico y las que tienen servicio de alcantarillado y así sucesivamente.

Gráfico 1. Matriz de correlaciones variables a fin del DDPCSP



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Aproximaciones del indicador DDPCSP: Elaboración de aproximaciones al indicador:

Debido a que hasta el momento los datos muestran que una manzana puede o no tener un alto volumen de manzanas que tienen servicios públicos y otros que no, se desea conocer para todo el conjunto de variables una variable representativa para cada tipo de población. Se entiende que la población favorecida refiere a la idea que son todas aquellas cantidades de viviendas que manifiestan tener algún tipo de servicio público y las no favorecidas son aquellas que no cuentan con ningún tipo de servicio público, es posible considerar la idea de agrupar estas dos poblaciones en 2 variables representativas del total de poblaciones. En ese orden de ideas, es necesario determinar qué operaciones es la que mejor condiciones estadísticamente aceptable permiten representar estas poblaciones, por lo que se crean dos posibles operaciones fiables a la hora de agruparlas, las cuales serían evaluadas para la selección solo de una a través de test estadísticos. Las dos operaciones son las siguientes:

Primera aproximación de la medición del indicador “DDPCSP”:

La aproximación con operación sumatoria de todas las poblaciones favorecidas:

Ecuación 1. Operación suma de la primera aproximación (PF) del indicador DDPCSP

Sumatoria Población Favorecida "SPF" = \sum Conteo de viviendas que cuentan con servicios

La segunda operación es la sumatoria de todas las poblaciones desfavorecidas, es decir que no cuentan con servicios públicos.

Ecuación 2. Operación suma de la primera aproximación (PNF) del indicador DDPCSP

Sumatoria Pobl. no Favorecida "SPNF" = \sum Conteo de viviendas que no cuentan con servicios

Se calcula el total:

Ecuación 3. Operación suma de la primera aproximación (TPA1) del indicador DDPCSP

$$\text{Total de población Aprox1TPA1} = SPF + SPNF$$

Ajustando la escala de los valores con la proporción para cada tipo de sumatoria con la siguiente ecuación:

Ecuación 4. Rescalado de la primera aproximación (SPF & SPNF)

$$\text{Sumatoria Población Favorecida "SPF" = } SPF/TPA1$$

$$\text{Sumatoria Población Favorecida "SPF" = } SPNF/TPA1$$

En la **primera aproximación** de la medición de este indicador y dada la sumatoria de las variables que indican favorecimiento debido a que la manzana registrada cuenta con un conteo de viviendas de algún tipo de servicio público mayor que cero, se establece una unidad de medida único de viviendas favorecidas con la división o fracción propiamente vista en la **Ecuación 4**, extendiéndose la misma concepción para las no favorecidas.

Segunda aproximación de la medición del indicador “DDPCSP”:

La aproximación con operación promedio de todas las poblaciones favorecidas se calcula con la siguiente ecuación:

Ecuación 5. Operación promedio de la segunda aproximación (PF) del indicador DDPCSP

$$\text{Promedio Población Favorecida "PPF" = } \frac{\sum \text{Conteo de viviendas que cuentan con servicios}}{\text{Cantidad de variables observadas}}$$

La segunda operación es el promedio de todas las poblaciones desfavorecidas, es decir que no cuentan con servicios públicos, calculado con la siguiente ecuación:

Ecuación 6. Operación promedio de la segunda aproximación (PPNF) del indicador DDPCSP

$$\text{Promedio Población no Favorec. "PPNF"} = \frac{\sum \text{Conteo de viviendas que no cuentan con serv.}}{\text{Cantidad de variables observadas}}$$

Se obtiene el total con la siguiente ecuación:

Ecuación 7. Operación suma de la segunda aproximación (TPA2) del indicador DDPCSP

$$\text{Total de población Aprox1TPA2} = PPF + PPNF$$

Se ajusta la escala de los valores con la proporción para cada tipo de sumatoria con las siguientes ecuaciones:

Ecuación 8. Rescalado de la segunda aproximación (PPF & PPNF)

$$\text{Sumatoria Población Favorecida "PPF"} = PPF / TPA2$$

$$\text{Sumatoria Población Favorecida "PPF"} = PPNF / TPA2$$

En la **segunda aproximación** de la medición de este indicador, se relaciona con el promedio de las variables que indican favorecimiento sobre las variables registradas, permitiendo establecer una posible unidad de medida única de viviendas favorecidas.

Tabla 7. Resumen estadístico de aproximaciones afines para la construcción del indicador en estudio (DDPCSP).

Poblacion_fav_aprox_1	Poblacion_nofav_aprox_1	Poblacion_fav_aprox_2	Poblacion_nofav_aprox_2
Min. : 0.0000	Min. : 0.00000	Min. : 0.0000	Min. : 0.0000
1st Qu.: 0.8960	1st Qu.: 0.02083	1st Qu.: 0.8295	1st Qu.: 0.1341
Median : 0.9362	Median : 0.05072	Median : 0.8471	Median : 0.1472
Mean : 0.8237	Mean : 0.06228	Mean : 0.7480	Mean : 0.1380
3rd Qu.: 0.9649	3rd Qu.: 0.08140	3rd Qu.: 0.8596	3rd Qu.: 0.1606
Max. : 1.0000	Max. : 1.00000	Max. : 0.8750	Max. : 0.5625
Poblacion_fav_aprox_2	Poblacion_nofav_aprox_2	Total_poblacion_aprox_2	
Min. : 0.0000	Min. : 0.0000	Min. : 0.00	
1st Qu.: 0.8295	1st Qu.: 0.1341	1st Qu.: 27.43	
Median : 0.8471	Median : 0.1472	Median : 59.43	
Mean : 0.7480	Mean : 0.1380	Mean : 92.07	
3rd Qu.: 0.8596	3rd Qu.: 0.1606	3rd Qu.: 102.29	
Max. : 0.8750	Max. : 0.5625	Max. : 4870.86	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE

En la **Tabla 7**, se muestra el resumen estadístico de estas cuatro variables creadas presentando diferentes comportamientos y distribuciones estadísticas, es posible identificar el significado de los resultados, entendiéndose que para la primera aproximación (operación suma) los máximos valores registrados para ambas poblaciones (favorecidas y no favorecidas) es el valor 1 (uno),

esto hace referencia que la sumatoria de todas las variables de tipo conteo de viviendas que cuentan con servicios existentes en una manzana cuentan con el 100% de los servicios (6 servicios en total).

Agregando una extensión conforme a las proporciones que las variables de favorecimiento presentan en función de las desfavorecidas, se suman los valores totales por variables calculadas, respectivamente:

Ilustración 1. Resultado operativo de sumatoria por columna de cada aproximación a fin para la construcción del indicador en estudio (DDPCSP).

Poblacion_fav_aprox_1	Poblacion_nofav_aprox_1	Poblacion_fav_aprox_2	Poblacion_nofav_aprox_2
35753.340	2703.660	32467.024	5989.976

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE

Empleando las siguientes operaciones para cada aproximación, se muestra los resultados.

Aproximación 1 (Operación sumatoria):

Ecuación 9. Operación fracción de la primera aproximación (SPF vs SPNF) del indicador DDPCSP

$$\text{Fracción de SPF vs SPNF} = \frac{\text{SPF}}{\text{SPNF}}$$

Los resultados muestran que la proporción de la sumatoria porcentual de viviendas con servicio de redes conectados es 13.22 veces mayor que la de sin servicios conectados(no favorecidas), entiéndase “servicio de redes” la alusión a todos los servicios públicos (servicio de gas, alcantarillado, etc), esto es que por cada unidad de manzana sin servicios existen 13 que si cuentan con algún tipo de servicio.

Imagen 6. Resultado de fracción de la primera aproximación (SPF vs SPNF) del indicador DDPCSP

$$(35753) / (2703)$$

$$13.22716$$

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE

Aproximación 2 (Operación promedio):

Ecuación 10. Operación fracción de la segunda aproximación (PPF vs PPNF) del indicador DDPCSP

$$\text{Fracción de PPF vs PPNF} = \frac{\text{PPF}}{\text{PPNF}}$$

En promedio por manzanas porcentualmente, la cantidad de viviendas favorecidas es 5.42 veces mayor que la de sin servicios conectados (véase la siguiente ilustración), es decir que por cada manzana en la ciudad que no cuenten con servicios existen en promedio 5.42 que si cuentan con servicios.

Imagen 7. Resultado de fracción de la segunda aproximación (PPF vs PPNF) del indicador DDPCSP

(32467) / (5989)

5.421105

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE

La fracción entre población favorecida vs no favorecida según la aproximación 2 es menor que la fracción de la operación suma(aproximación 1), lo que es equivalente que a diferencia de la aproximación 1 la operación promedio regula mejor los Outliers o cifras atípicas encontradas en los registros de la base de datos.

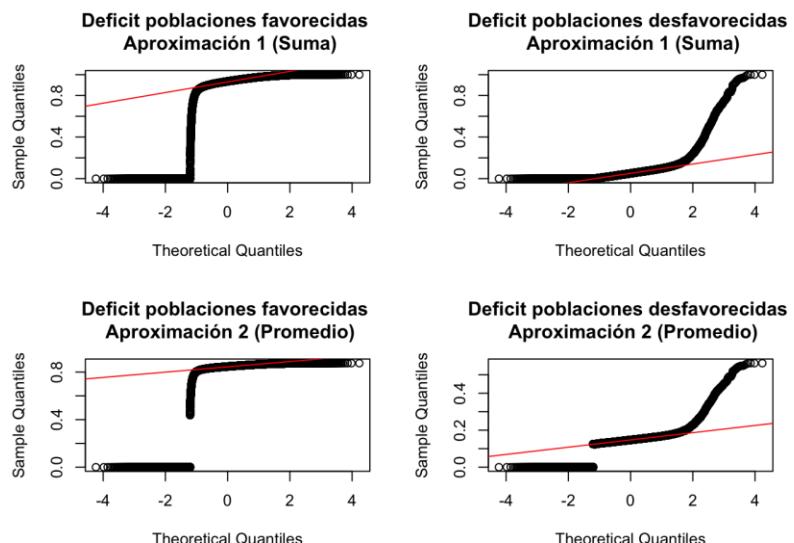
Con estos resultados se procede entonces a seleccionar para ambas operaciones, la que mejor podría explicar el déficit teniendo en cuenta que hacen referencias a valores estáticos en el tiempo y/o de muestras que buscan explicar las condiciones habitacionales de cobertura en servicios públicos, por tanto en primera medida se procede a aplicar un análisis de inferencia estadística para encontrar percepciones única, propias de esta distribución de datos.

Análisis estadísticos de las aproximaciones DDPCP:

Verificación de condiciones de Normalidad para ambas aproximaciones:

Se busca determinar bajo los supuestos de normalidad el cumplimiento de condiciones de distribución normal para proceder a aplicar tests estadísticos que permitan medir las estimaciones poblaciones en cuanto a los indicadores, favoreciendo al entendimiento de las condiciones de distribución de los datos y tomar la decisión de elegir sobre el tipo de operación con mejores condiciones estadísticas, por tanto es necesario estudiar las condiciones de normalidad de las aproximaciones.

Gráfico 2. Gráficos Q-Q para cada aproximación del indicador DDPCP



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

En el **Gráfico 2** de normalidad, se observa que para los indicadores de "Déficit poblaciones favorecidas" (Suma & Promedio), existe un detallada diferencia entre la forma de las graficas de QQ(cuartiles por cuartiles), tanto así que los valores para la aplicación de la operación promedio resultan con menos continuidad comparado a la suma, a diferencia de las variables de "Déficit poblaciones desfavorecidas", que sus gráficos permanecen con un tanto idénticos, no obstante para todas las variables se encuentran que no cumplen el supuesto de normalidad debido a que no se aproximan en forma lineal a la línea de tendencia.

Aplicación de prueba Kolmogorov- Smirnov de normalidad:

Debido a que los supuestos de normalidad anteriormente planteados, no resultan suficientes para explicar la proveniencia de los datos y que estos cumplan criterios de normalidad, se procede a aplicar la prueba Kolmogorov- Smirnov para cada variable, de forma que permita evaluar con grado de significancia las condiciones de normalidad de los datos:

Tabla 8. Resultado de aplicación de test estadísticos de Normalidad Bilateral.

```

One-sample Kolmogorov-Smirnov test          One-sample Kolmogorov-Smirnov test

data: df_3$Poblacion_fav_aprox_1           data: df_3$Poblacion_nofav_aprox_1
D = 0.37333, p-value < 2.2e-16            D = 0.19793, p-value < 2.2e-16
alternative hypothesis: two-sided          alternative hypothesis: two-sided

One-sample Kolmogorov-Smirnov test          One-sample Kolmogorov-Smirnov test

data: df_3$Poblacion_fav_aprox_2           data: df_3$Poblacion_nofav_aprox_2
D = 0.42556, p-value < 2.2e-16            D = 0.2976, p-value < 2.2e-16
alternative hypothesis: two-sided          alternative hypothesis: two-sided

```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Resultado de la prueba de Normalidad:

Para las cuatro variables de déficit entorno, correspondientes a la operación de suma de poblaciones favorecidas y no favorecidas adicionalmente la operación promedias de estas variables, es posible afirmar que se puede descartar la posibilidad que todas variables tengan una distribución normal con el 95% de confiabilidad, trayendo consecuencias estadísticas, tales como:

- i) Los estimadores para optimizar los mínimos cuadrados para asumir el mejor valor en términos de la mínima varianza no resultan eficientes a la hora de tratar estas variables.
- ii) Las posibles estimaciones estadísticas de estas variables en términos de intervalos que permiten aceptar una aproximación teórica de comportamiento estadístico paramétrico y de contraste de significancia resultan solo aproximados e inexactos.

Al llegar esta conclusión, se puede extender a analizar la cuantificación conforme a la distancia entre distribuciones de variables, en este caso seleccionar la mejor operación para dar continuidad a la medición del indicador y también para determinar en el caso de que sean iguales para alguna, si no es de peso estadístico tener preferencias por una variable en específico.

(Rodrigo, 2021)

Diferencia estadísticamente significativa entre las aproximaciones del indicador:

Para establecer la diferencia estadísticamente significativa entre las dos aproximaciones, en primera instancia se calcula los índices de homocesdasticidad es decir la determinación de homogeneidad de la varianza, aplicando test estadísticos no paramétricos.

Soporte : Test de Fligner-Killeen

Este test no paramétrico permite comparar las variancias con base a la mediana, por tanto buscamos con este explicar la diferencia entre las dos aproximaciones de las variables (sumas y promedios), capaces de calcular la probabilidad exacta de observar diferencias entre distribuciones superiores que otras y/o si en realidad existiese un diferencia significativa, para

poblaciones en las que se ha comprobado que no se distribuyen normalmente, respectivamente:

Ilustración 2. Resultado de aplicación de test estadístico Fligner-Killeen para evaluación de Homocesdasticidad entre representaciones de poblaciones (PF vs PNF).

```
Fligner-Killeen test of homogeneity of variances

data: list(df_3$Poblacion_fav_aprox_1, df_3$Poblacion_nofav_aprox_1)
Fligner-Killeen:med chi-squared = 1336.5, df = 1, p-value < 2.2e-16

Fligner-Killeen test of homogeneity of variances

data: list(df_3$Poblacion_fav_aprox_2, df_3$Poblacion_nofav_aprox_2)
Fligner-Killeen:med chi-squared = 420.71, df = 1, p-value < 2.2e-16
```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Debido a el p-valor para ambos test, referentes a la comparativa de varianza entre las aproximaciones son menores que 0.05 (grado de significancia), se puede aceptar la idea que el conteo de viviendas que tienen servicios de redes tiene diferencias en cuanto a la variabilidad con respecto a las que no tienen servicio, además que el estadístico chi-squared calculado para ambos test demuestra que la “operación de promedio de valores”, tiene menores discrepancias entre las frecuencias obtenidas, por lo que tiene mayor concordancia perfecta entre las frecuencias observadas y las esperadas.

(Mitjana, 2021)

Aproximaciones absolutas del indicador DDPCSP:

Una vez construido y evaluado lo que serían los insumos para la construcción definitiva del indicador, se procede entonces a realizar las siguientes operaciones para cada conjunto de variable.

Aproximación absoluta 1(Operación suma):

Se obtiene a partir de la diferencia entre la sumatoria de población favorecida contra la desfavorecida.

Al obtener el diferencial entre la población que cuenta con servicios de redes y las que no cuentan con servicios, se estaría midiendo las proporciones que predominan en cada registro de manzana, permitiendo validar en que zonas poblacionales existe mayoría de población que cuenta con servicios. Por lo anterior, es posible tomar la decisión de cual variable permitiría explicar con menor variabilidad el indicador de déficit habitacional de entorno en referencia a la aplicación de estadísticos no paramétricos, en esa dirección se estaría obteniendo el indicador que plantea la siguiente idea: Sí se reduce la población de viviendas para una

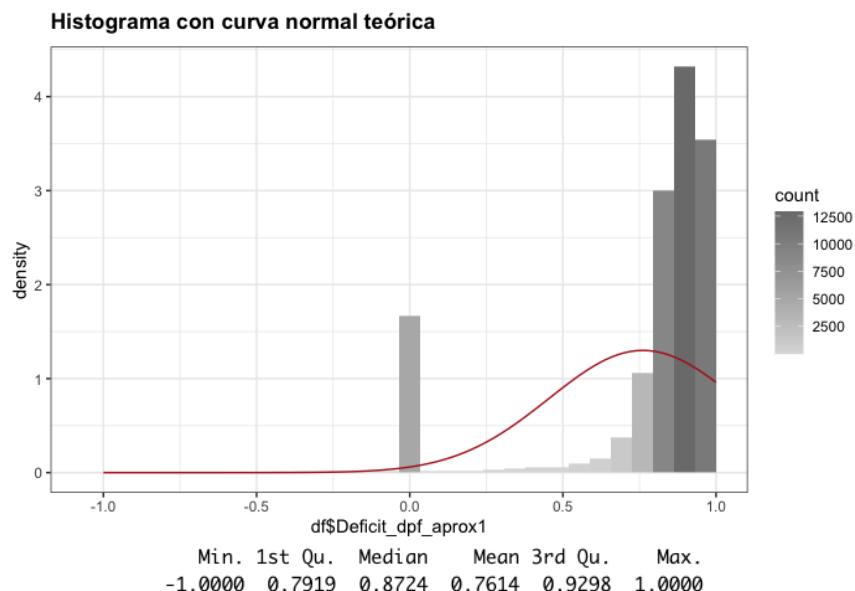
manzana que cuentan con servicio contra las que no cuenta con servicio, ¿Cuál es el resultado?, en caso de ser positivo es porque hay un superávit de población que cuenta con servicios y en caso de ser negativo es que existen un déficit de población que cuenta con servicios o un superávit de población desfavorecida.

Ecuación 11. Aproximación absoluta del indicador déficit diferencial de población favorecida (DDPCSP).

$$\text{Deficit diferencial PF (Aproximación 1)} = [\text{SPF} - \text{SPNF}]$$

Ejecutando esta aproximación absoluta la distribución de los datos muestra un asentamiento de curva con sesgo a la izquierda (asimetría), se observa en el histograma que hay una gran concentración de datos ($12500 + 10000 + 7500 = 30000$) +30000 manzanas para el valores mayores que 0.75 del indicador lo que hace referencia que la población en la mayor cantidad de manzanas cuentan con servicios y el excedente de la población restante no cuentan con servicios.

Gráfico 3. Histograma con curva normal teórica de la primera aproximación absoluta del DDPCSP.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

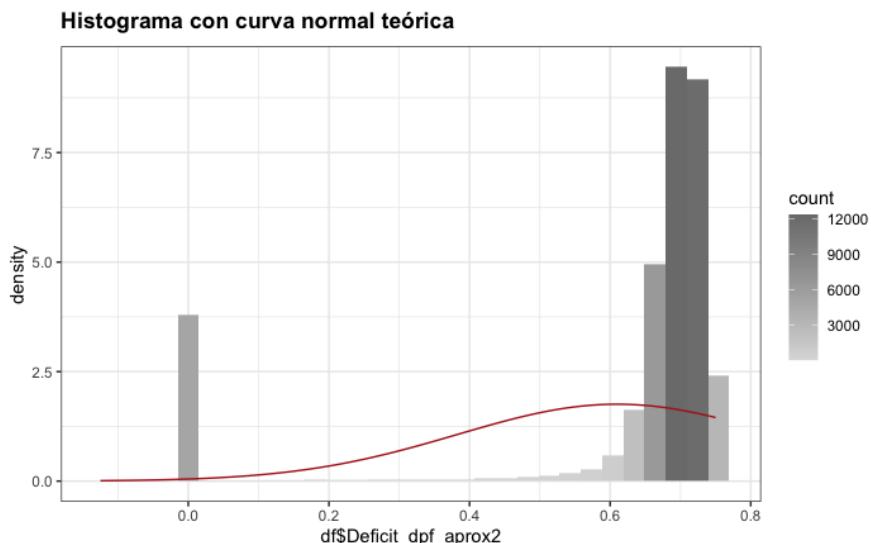
La curva normal teórica muestra que a pesar que la gran cantidad llega a su máximo antes del valor 1 para el indicador, se presenta un descenso de niveles de indicador en cuanto a la densidad de manzanas que se ubican con esa magnitud del indicador, y que esta curva es leptocúrtica lo que indica que los datos tienen menor dispersión.

Aproximación absoluta 2(Operación promedio):

Se obtiene a partir de la diferencia entre el promedio de población favorecida contra la desfavorecida.

Ecuación 12. Aproximación absoluta del indicador déficit diferencial de población favorecida (DDPCSP).

$$\text{Deficit diferencial PF (Aproximación 2)} = [\text{PPF} - \text{PPNF}]$$

Gráfico 4. Histograma con curva normal teórica de la segunda aproximación absoluta del DDPCSP.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

La distribución de los datos en la aproximación absoluta 2 del indicador es asimétrica, presenta un sesgo hacia la izquierda, esto presentando unas componentes del indicador de menor uniformidad para las distribuciones en cada manzana, debido a que su curva teórica muestra que los datos no presentan tantos valores máximos o de menor incertidumbre, esto es debido a que la curva es de tipo platicúrtica lo que infiere que existe una dispersión alta de valores distribuidos a lo largo del data set.

Selección de la mejor aproximación conforme al Test de varianza:

Test de Levene: Estadísticos No paramétricos

Aplicando el **test de Levene** caracterizado por comparar 2 (dos) o más poblaciones conforme a sus diferencias significativas entre varianzas y compilando la variable categórica correspondiente a las localidades, se aplica el test para cada diferencial de poblaciones de las aproximaciones expuestas anteriormente:

Ilustración 3. Resumen de aplicación test Levene

Homogeneidad de varianza del indicador por cada localidad.

```
Levene's Test for Homogeneity of Variance (center = "median")
  Df F value    Pr(>F)
group   18  30.98 < 2.2e-16 ***
43312
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Levene's Test for Homogeneity of Variance (center = "median")
  Df F value    Pr(>F)
group   18  21.154 < 2.2e-16 ***
43312
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El test encuentra diferencias significativas entre las varianzas evidentemente de las localidades de la ciudad, y a partir del valor de "Fisher, la primera aproximación absoluta muestra un "F-Value" es mayor a el diferencial de población con la segunda aproximación absoluta, esto indica que las distribuciones del indicador (DDPCSP) por localidad del tipo operación promedio son más uniformes comparado con la aproximación absoluta de la operación suma.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Resumen de la determinación de la mejor aproximación al indicador:

El resumen estadístico que demuestra el ajuste de este indicador, se presenta a continuación:

Tabla 9. Consolidado de estadísticos aplicados a cada aproximación absoluta para el indicador DDPCSP

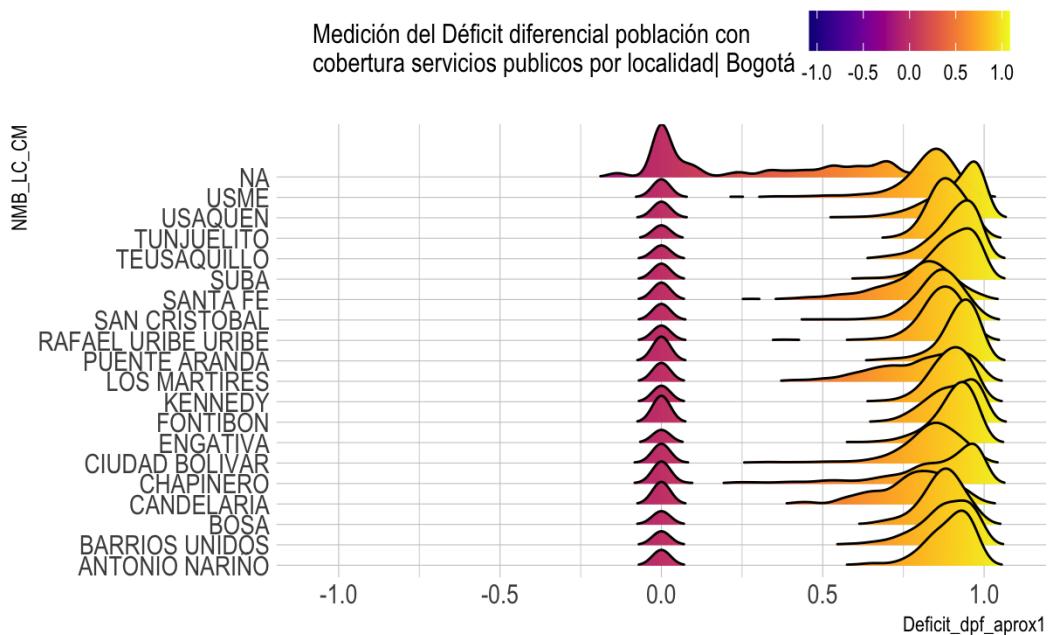
Aproximación	Estadístico Fligner-Killeen	Resumen estadístico	Estadístico Levene
Aprox. Abs 1 (Suma)	Alta significancia y Alta discrepancia entre las frecuencias obtenidas, tiene menor concordancia perfecta entre las frecuencias observadas y las esperadas.	Curva normal teórica leptocúrtica: Menor Dispersión	Alta significancia y Menor uniformidad de las distribuciones por localidades.
Aprox.Abs 2 (Promedio)	Alta significancia y Baja discrepancia entre las frecuencias obtenidas, tiene mayor concordancia perfecta entre las frecuencias observadas y las esperadas.	Curva normal teórica Platicúrtica: Mayor dispersión.	Alta significancia y Mayor uniformidad de las distribuciones por localidades.

Finalmente, el consolidado comparativo de las aproximaciones permiten afirmar que para el indicador, debe optarse por la selección de la aproximación absoluta 1 (operación suma), debido a que este cuenta con altas discrepancias de valores más sujetos a la realidad, esto debido a que las poblaciones se distribuyen de forma aleatoria además de presentar muchas diferencias entre sí, por conseciente contando con un sesgo cuya distribución de los datos presentan baja dispersión y una uniformidad promedio en la distribución por localidades debido a que a pesar que las varianzas no son homogéneas para las localidades su semejanza relativa en la distribución de los datos por cada localidad es aceptable. Adicionalmente se demuestra que la distribución de los datos están concentrados para unas medidas de tendencias central sin extremos diferenciales para la media y la mediana, que a su vez el indicador de déficit diferencial de poblaciones con cobertura a servicios públicos (DDPCSP) refiere de ser proveniente de la operación suma, con estas evidencias se justifica con un alto potencial para constatar una contundente explicación de la realidad de las poblaciones favorecidas en los niveles de proporción de los datos estimulando la plausibilidad de la incidencia de las proporciones poblacionales.

(Rodrigo, Análisis de la Homogenidad de varianza (Homocesdasticidad), 2021)

Categorización del indicador DDPCSP:

Gráfico 5. Gráfico de densidad del indicador DDPCSP por localidad en la ciudad de Bogotá DC.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Los gráficos de densidad por localidad mostrados en el **Gráfico 5**, permiten apreciar que gran cantidad de datos están ubicados en la mediana de la distribución de los datos, de lo cual es posible inferir que aquella distribución que no tenga similitud con la general teórica presentan condiciones especiales, tal como lo presentan la población de localidad que no cuenta con registro (NA), cuya cantidad de datos se concentran en niveles de igualdad a cero del indicador, esto es que dentro de la escala la cantidad de población que no cuenta con servicios es la misma cantidad de población que si cuenta con servicios. Se entiende que las manzanas que se encuentren por debajo de 0 serán consideradas en déficit, es decir existe mayor cantidad de viviendas que no cuentan con servicios públicos que las que si, las que se encuentren entre los rangos del 0 al 1 se afirma que están por encima del promedio y por lo tanto en superávit, esto indica que existe mayor población de viviendas que si cuentan con servicios públicos.

Las categorías discriminadas en rangos equivalentes para establecer un tipo de déficit o superávit son las siguientes:

- $[-\infty, -1] \rightarrow "Crítico Déficit"$
- $[-1, -0.75] \rightarrow "Alto Déficit"$
- $[-0.75, -0.5] \rightarrow "Moderado Déficit"$
- $[-0.5, -0.25] \rightarrow "Medio Déficit"$
- $[-0.25, -0.05] \rightarrow "Déficit"$
- $[-0.05, 0.05] \rightarrow "Equilibrio"$
- $[0.05, 0.25] \rightarrow "Superávit"$
- $[0.25, 0.5] \rightarrow "Medio Superávit"$
- $[0.5, 0.75] \rightarrow "Moderado Superávit"$
- $[0.75, 0.99] \rightarrow "Alto Superávit"$
- $[0.99, +\infty] \rightarrow "Crítico Superávit"$

Para el valor de uno (+1), indica que el 100% de las viviendas observadas en esa manzana de estudio cuentan con cobertura a servicios públicos y en un caso extremo un valor del indicador igual a -1 (menos uno) estaría contemplando que en esa manzana el 100% de la población no cuenta con cobertura a servicios públicos.

Análisis a partir de la tabla de contingencia del indicador DDPCSP

La tabla de contingencia permite realizar el cruce de información entre las categorías y las localidades. En este caso se acude a un análisis descriptivo del indicador con respecto a las localidades de la ciudad, teniendo en cuenta el promedio de indicador por localidad, la tabla obtenida es la presentada en la **Tabla 10**.

Tabla 10. Tabla de contingencia del indicador para cada localidad.

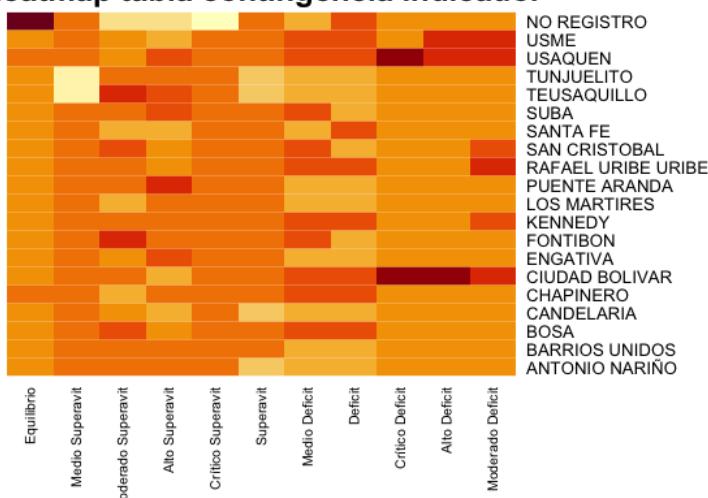
	Equilibrio <dbl>	Medio Superavit <dbl>	Moderado Superavit <dbl>	Alto Superavit <dbl>	Crítico Superavit <dbl>	Superavit <dbl>	Medio Deficit <dbl>	Deficit <dbl>	Crítico Deficit <dbl>	Alto Deficit <dbl>	Moderado Deficit <dbl>
ANTONIO NARIÑO	0.000000e+00	0.444444	0.6816952	0.8978519	0.9977002	-1.0000000	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
BARRIOS UNIDOS	0.000000e+00	0.4583333	0.6897371	0.8923668	0.9969949	0.08333333	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
BOSA	-9.57960e-05	0.3880099	0.6957724	0.8785468	0.9983248	0.1858839	-0.3379884	-0.1246828	-1.00	-1.0000000	-1.0000000
CANDELARIA	0.000000e+00	0.4422461	0.6691105	0.8476971	1.0000000	-1.0000000	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
CHAPINERO	3.227537e-04	0.4039406	0.6563472	0.8983394	0.9970739	0.17047995	-0.2533937	-0.1367019	-1.00	-1.0000000	-1.0000000
CIUDAD BOLÍVAR	-2.881911e-04	0.3875247	0.6840334	0.8617319	0.9983354	0.14682500	-0.3538125	-0.1399802	-0.99	-0.8760665	-0.6038994
ENGATIVA	0.000000e+00	0.3958268	0.6791268	0.9079903	0.9975163	0.25000000	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
FONTIBÓN	0.000000e+00	0.4149360	0.7117416	0.8995594	0.9971801	0.25000000	-0.2720000	-1.0000000	-1.00	-1.0000000	-1.0000000
KENNEDY	1.107972e-04	0.4147757	0.6869689	0.8963459	0.9980341	0.11421228	-0.3636364	-0.1270972	-1.00	-1.0000000	-0.6666667
LOS MARTIRES	0.000000e+00	0.4482531	0.6617277	0.8835012	0.9980433	0.2222222	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
PUENTE ARANDA	0.000000e+00	0.4047619	0.6824297	0.9209965	0.9978633	0.14285714	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
RAFAEL URIBE URIBE	-1.271876e-05	0.3808731	0.6921633	0.8754758	0.9992841	0.17185800	-0.3819444	-0.1139601	-1.00	-1.0000000	-0.6200000
SAN CRISTOBAL	6.773101e-06	0.4211233	0.6982033	0.8697571	0.9996333	0.14796528	-0.2612613	-1.0000000	-1.00	-1.0000000	-0.6666667
SANTA FE	4.304964e-05	0.4148000	0.6629595	0.8479774	1.0000000	0.20009158	-1.0000000	-0.2444444	-1.00	-1.0000000	-1.0000000
SUBA	0.000000e+00	0.4121901	0.6803736	0.9054107	0.9974553	0.1444444	-0.3245614	-1.0000000	-1.00	-1.0000000	-1.0000000
TEUSAQUILLO	0.000000e+00	-1.0000000	0.7076399	0.9114490	0.9959716	-1.0000000	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
TUNJUELITO	0.000000e+00	-1.0000000	0.6910131	0.8859392	1.0000000	-1.0000000	-1.0000000	-1.0000000	-1.00	-1.0000000	-1.0000000
USAQUÉN	2.954206e-04	0.3869812	0.6765082	0.9133976	0.9967865	0.14364858	-0.2668011	-0.1354839	-0.99	-0.9375000	-0.5081967
USME	-2.742371e-04	0.4005015	0.6747940	0.8524992	0.9996047	0.17174847	-0.3923121	-0.1375693	-1.00	-0.9166667	-0.5778444
NO REGISTRO	3.943543e-03	0.4093073	0.63337987	0.8275132	-1.0000000	0.13320537	-1.0000000	-0.1368421	-1.00	-1.0000000	-1.0000000

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

En total son 11 categorías que buscan explicar en cada localidad como es el comportamiento del indicador, teniendo en cuenta que la operación de la creación de la tabla de contingencia (tabla cruzada), es el promedio del indicador para cada de registro en la localidad, esto es posible apreciarlo en un mapa de calor en los que los valores se muestran con respecto a una escala por tipo de categoría, el que a mayor oscuridad mayor es la magnitud.

Ilustración 4. Mapa de calor de la tabla de contingencia del indicador DDPCSP

Heatmap tabla contingencia indicador



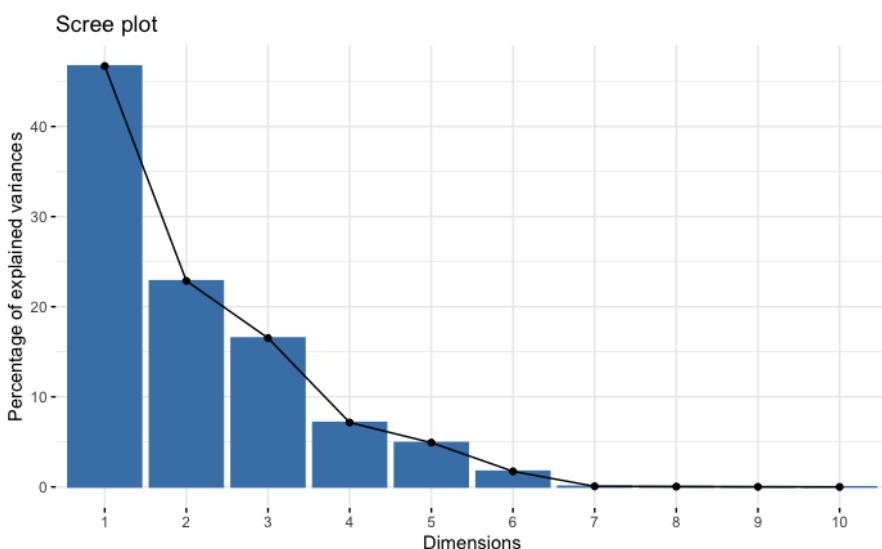
En la localidad de Ciudad Bolívar y Usaquén en promedio se presentan manzanas con niveles Críticos y medio déficit, la localidad de Usme presenta en promedio altas concentraciones de niveles en Déficit Medio. Los niveles de Alto Superávit se destacan para las localidades Puente Aranda, Engativá, Teusaquillo, Suba y Usaquén. Finalmente, se presentan alta magnitud del indicador en promedio para manzanas con Moderado Superávit del índice diferencial de poblaciones con cobertura a servicios públicos, en la

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

localidad de Teusaquillo, se destaca junto con la localidad de Fontibón en esa categoría . Es de notarse que para todas las localidades se presentan así mismo registros de manzanas de tipo Nivel Equilibrio y Crítico Superávit, es decir es normal encontrar población favorecidas con conexión a servicios públicos y las no favorecidas se equilibran en términos de cantidad.

Análisis de componentes principales (PCA) exploratorio para el indicador:

Gráfico 6. Scree plot: PCA gráfico de barras de componentes.



La técnica de análisis de componentes principales exploratorio para tablas de contingencias son esenciales a la hora de establecer clústeres e identificar causalidades entre los diferentes factores dentro de una información de resultado tal como se presenta la tabla cruzada, permitiendo cuales de estos factores comparten perfiles y de acuerdo a las categorías como es el comportamiento esperado y la dirección esperada

para cada factor. En este caso los factores hacen alusión a las localidades que están siendo medidas con el indicador Déficit diferencial de población con cobertura a servicios públicos.

La evaluación de las cantidades de componentes necesarias para explicar un porcentaje específico de la varianza de los datos se presenta gracias al "Scree plot", este indica que con dos componentes principales es posible explicar más del 70% (47% primer componente + 23% segunda componente) de la variabilidad de los datos y que el otro 30% de información es disipada.

Aplicando el método circular y gráfico PCA, asumiendo una perdida de información de exactamente 30.4% se obtiene los siguientes resultados:

Diagrama 2. PCA método gráfico de tablas

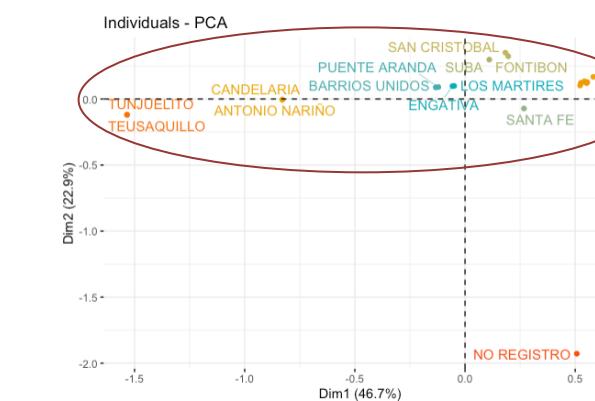
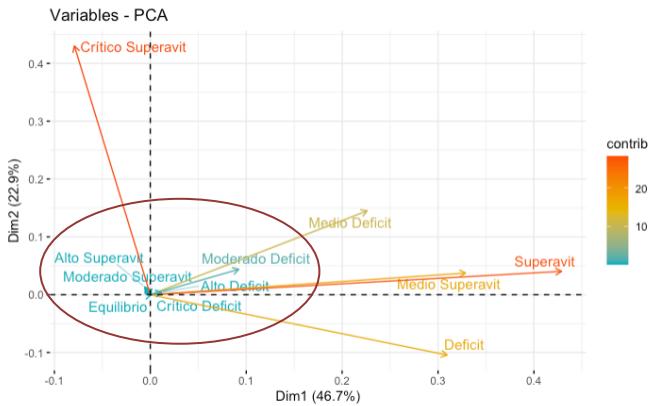


Diagrama 1. PCA método circular de tabla cruzada.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Los diagrama de PCA obtenidos, permiten ver los patrones de comportamiento de los indicadores en función de cada categoría del indicador, explicados en un porcentaje específico de la cantidad de datos a través de dos componentes principales. Se muestra, que para dos componentes principales que buscan explicar el comportamiento de la variabilidad en un 69.6% la mayor cantidad de localidades delimitadas por elipses de contorno rojos, se encuentran ubicadas muy próximas al origen indicando que presentan distribuciones de valores para el valor esperado del indicador y los mismos puntos débiles de los valores de indicador, esto es que no se diferencian unas con otras en la mayoría de atributos categóricos. La medición del indicador para estos factores presentan perfiles similares, es posible encontrar una asociación de valores por localidad que comparten distribuciones de medición del indicador intercorrelacionados.

Los colores encontrados en el análisis PCA, representan los perfiles de las localidades que más relación tienen con respecto a las categorías y como es el grado de contribución en función de la cantidad de localidades, para el caso del indicador DDPCSP, que las localidades cuentan con un perfil representativo en la categorías de un mismo color por ejemplo el color "0 azul" es la agrupación de distintas categorías para un conjunto de localidades con características particulares y diferentes, por ejemplo a la localidad con color 0 de Teusaquillo la cual presenta en promedio niveles del indicador a estar en dirección de "Crítico Superávit", además de encontrar la relación entre la clasificación de los valores en método gráfico que establece las

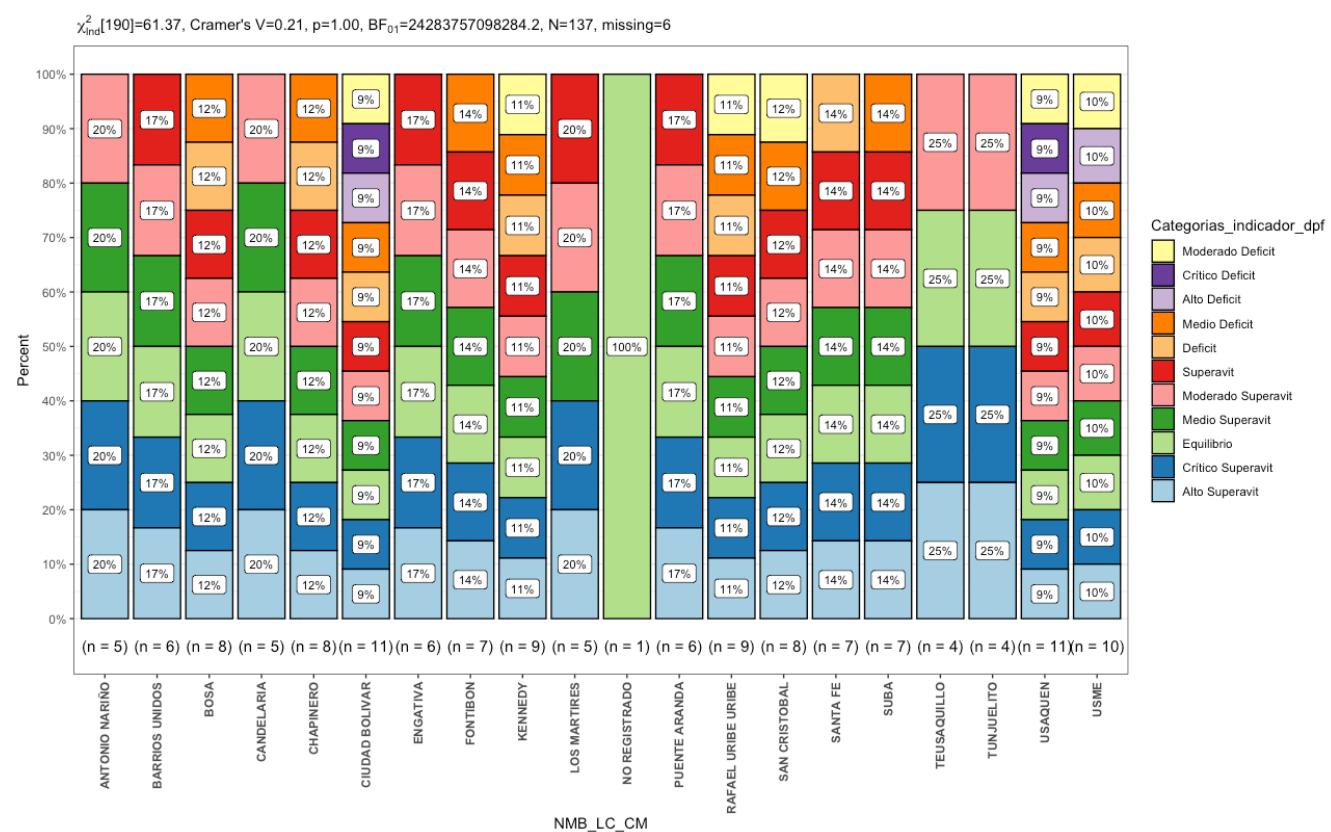
localidades con mayor ajuste relativo de acuerdo al perfil y los resultados de los valores en promedio del indicador por categoría.

(Chavez R. , 2020)

Los gráficos cruzados de barras permite visualizar de forma practica, los porcentajes de participación conforme a las manzanas registradas de la base de datos, respectivamente:

Análisis de tablas de contingencia para el indicador : "DDPCSP"

Gráfico 7. Gráfico de tabla cruzadas del indicador DDPCSP.

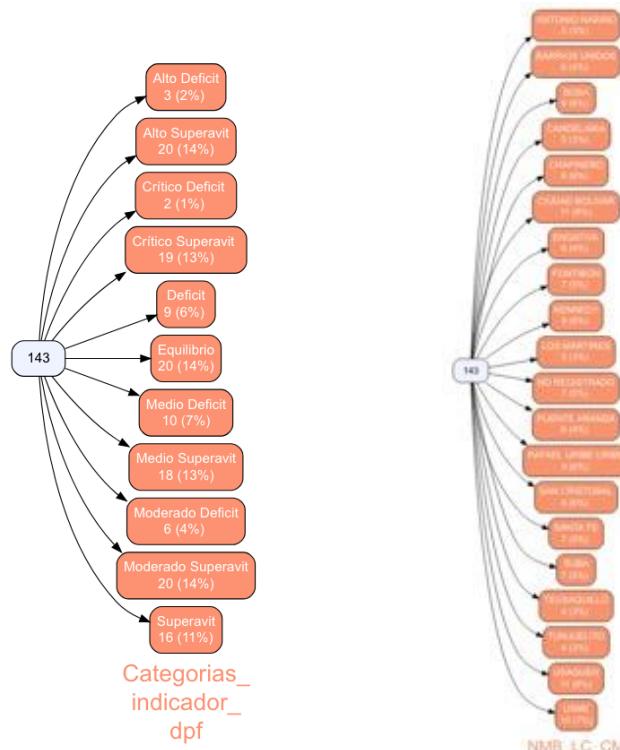


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

El anterior **Gráfico 7**, presenta las participaciones de las categorías por localidad, por tanto es de notarse que para las localidades de "Ciudad Bolívar" y "Usaquén" que 9 de cada 100 manzanas se encuentran en todas las categorías sin excepción, para las localidades de "Teusaquillo" y "Tunjuelito" estas con el mejor nivel del indicador expresan que uno de cada cuatro manzanas se encuentran en promedio en niveles del indicador de tipo " Alto Superávit", "Crítico Superávit", "Equilibrio" y "Moderado Superávit", para el caso de las localidades "Santa Fe" y "Suba" las participaciones de las categorías refieren de similitud a diferencia de una

categoría en la cual para la primera localidad esa participación se encuentra en "Déficit" y para la segunda localidad su participación en esa categoría diferencial es un nivel más elevado del déficit "Medio Déficit".

Diagrama 3. Diagrama de árbol de participación por categoría y por localidad.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

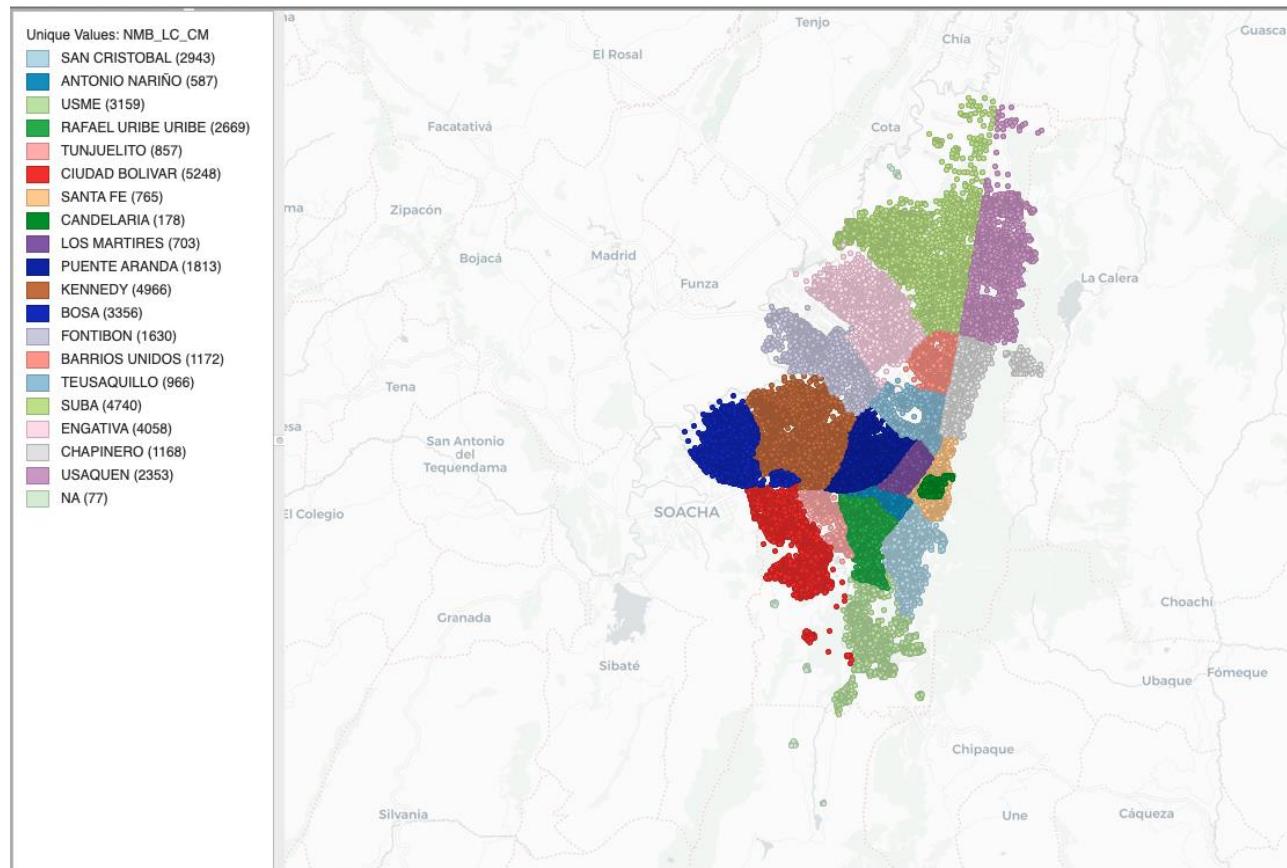
Los porcentajes para cada categoría del indicador en promedio por manzana agrupados por localidad varia gradualmente, para los 143 valores de los registros a priori del indicador preliminar a la tabla de contingencia (cruzada), la mayor concentración esta dada a nivel general por registros ubicados en las categorías: "Alto Superávit", "Equilibrio" y "Moderado Superávit" con un 14% para cada una, seguido de "Crítico Superávit" con un 13% y "Superávit" con un 11%, lo que equivale a decir aproximadamente que el 66% de las manzanas se encuentran en estas categorías y el 34% restantes se muestran en el diagrama . Las participaciones por localidad muestran que el valor esperado de participación por manzana se encuentre en 5%. (Chavez, 2021)

Construcción de mapas temáticos de la medición del indicador “DDPCSP”:

Análisis de localidades con mediciones elevadas del indicador :DDPCSP

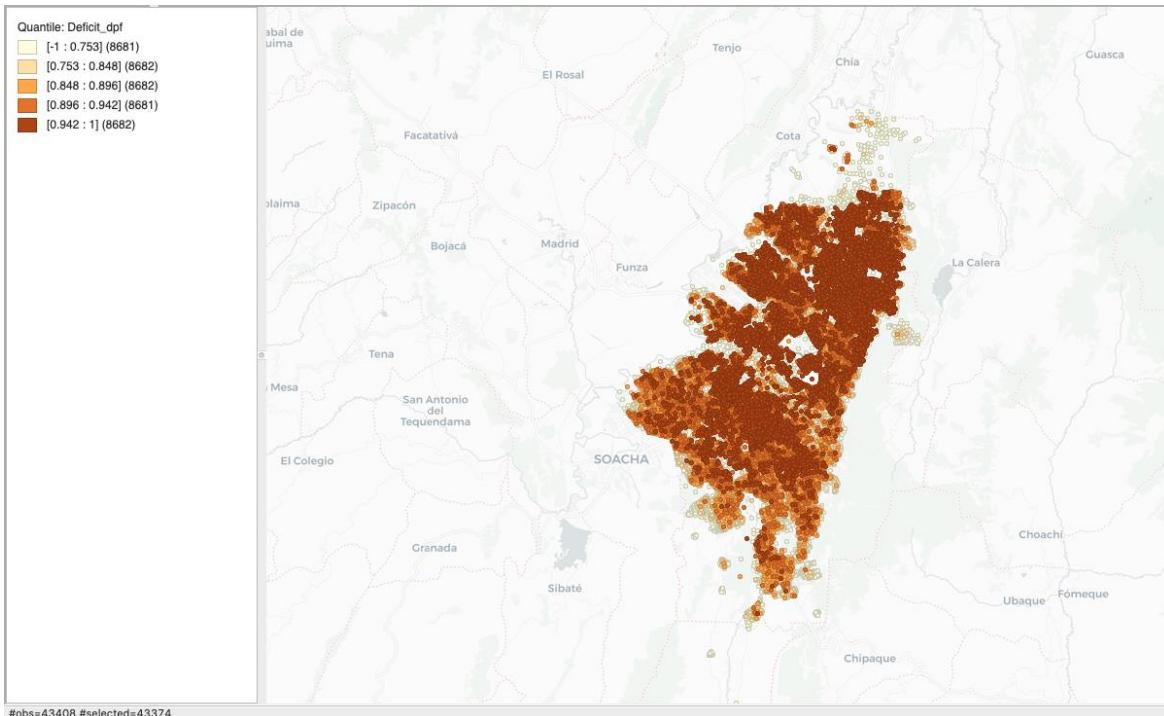
Se muestran en la siguiente visualización las manzanas por localidad, estas manzanas en un plano bidimensional delimitado por latitudes y longitudes. La localidad “Ciudad Bolívar” asciende al máximo valor de manzanas registradas, con un total de 5248 manzanas, comparado con sus localidades vecinas que son “Bosa” con un total de 3356 manzanas, “Tunjuelito” con solo 857 manzanas y con todas las localidades registradas.

Ilustración 5. Mapa distribución geoespacial de las manzanas (datos vectoriales) en la ciudad de Bogotá DC por localidad registrada (Gráfico geoespacial único de localidades).



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

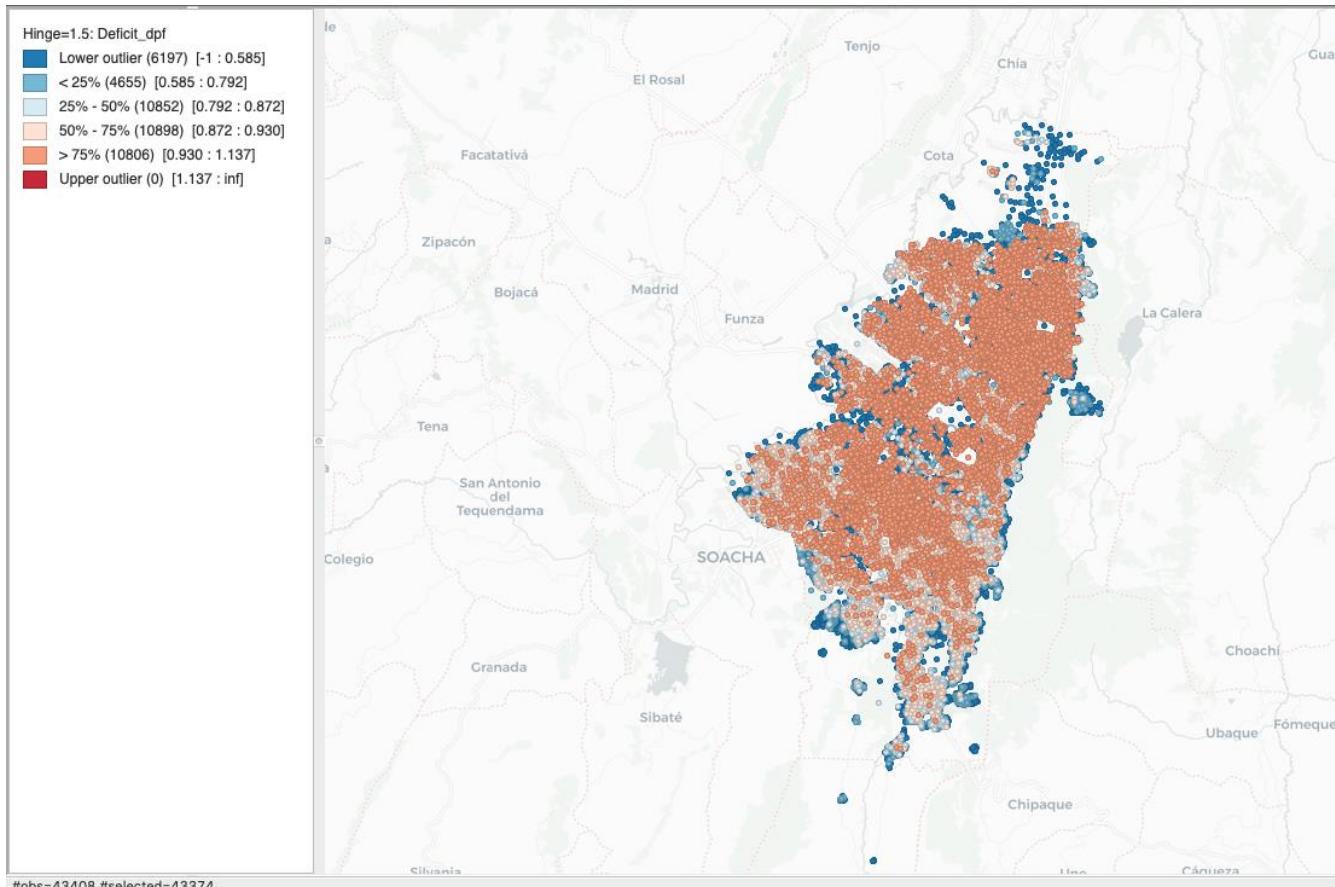
Ilustración 6. Mapa de distribución geoespacial del indicador DDPCSP en Cuartiles. (5)



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

Los mapas de cuartiles representan la clasificación del valor del indicador en la ciudad de Bogotá DC, para un volumen de manzanas iguales entre categorías de orden ascendente el superávit habitacional de entorno se pronuncia de forma uniforme en la ciudad de Bogotá, a excepción de la zona central. Una clasificación que contemple los rangos intercuartílico de la distribución de los datos, se muestra en la **Ilustración 7**, la presencia de Outliers, por encima y por debajo del rango intercuartílico. Las localidades donde se concentra mayor cantidad de manzanas con superávit se encuentran en el noreste de la ciudad.

Ilustración 7. Mapa de distribución geoespacial del indicador DDPCSP en Cuartiles. (1.5 Rango Intercuartílico). (Gráfico de clasificación intercuartílico)



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

Análisis espacial de los resultado de la medición del indicador “DDPCSP”:

Análisis de localidades con mediciones estadísticamente significativas del indicador :

Los HotSpots y ColdSpots con mediciones estadísticas que permiten responder a interrogantes

trascendentales en la distribución de los datos, así mismo disminuir la subjetividad con el respaldo de la visualización de los datos a través del soporte estadístico que clasifica a las zonas en factores de escala con mayor incidencia e influencia por los datos en estudio.

En este análisis, se toma la modelación de la auto correlación de los valores de cada manzana conforme a la vecindad a partir de un gestor de matriz de pesos por contigüidad, es

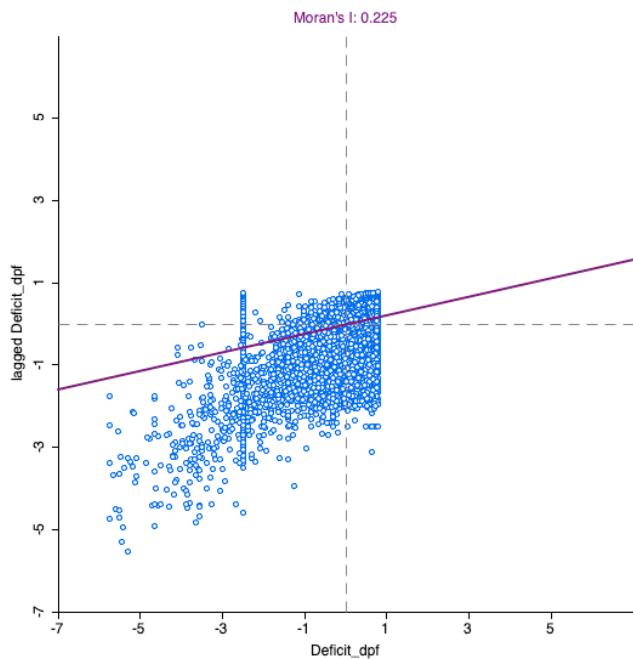
Property	Value
type	queen
symmetry	symmetric
file	df_m_1er_ind_MGN20211119.gal
id variable	field_1
order	1
# observations	43408
min neighbors	3
max neighbors	30
mean neighbors	6.00
median neighbors	6.00
% non-zero	0.01%

Tabla 11. Resultados de la creación de matriz de valores de pesos para el data set MGN-NIM 2018- Bogotá DC.

decir conforme a la distribución espacial y la superposición de estas manzanas se busca medir el grado de correlación que tienen una manzana con otras cercanas en el orden de 1, obteniéndose que el máximo numero de vecinos que cumplen con el criterio de contigüidad y que por ende puede formarse una cadena de pesos, donde el máximo número de vecinos que se pueden formar es 30 y que el mínimo es de 3, es decir una triada de manzanas que debido a su condiciones de proximidad podrían explicarse para un mismo parámetro estadístico, por consiguiente el promedio de los vecinos óptimos para un conjunto de peso es de 6.

Gráfico de dispersión de Morán:

Gráfico 8. Gráfico de dispersión del índice de Morán (DDIM) : DDPCSP.



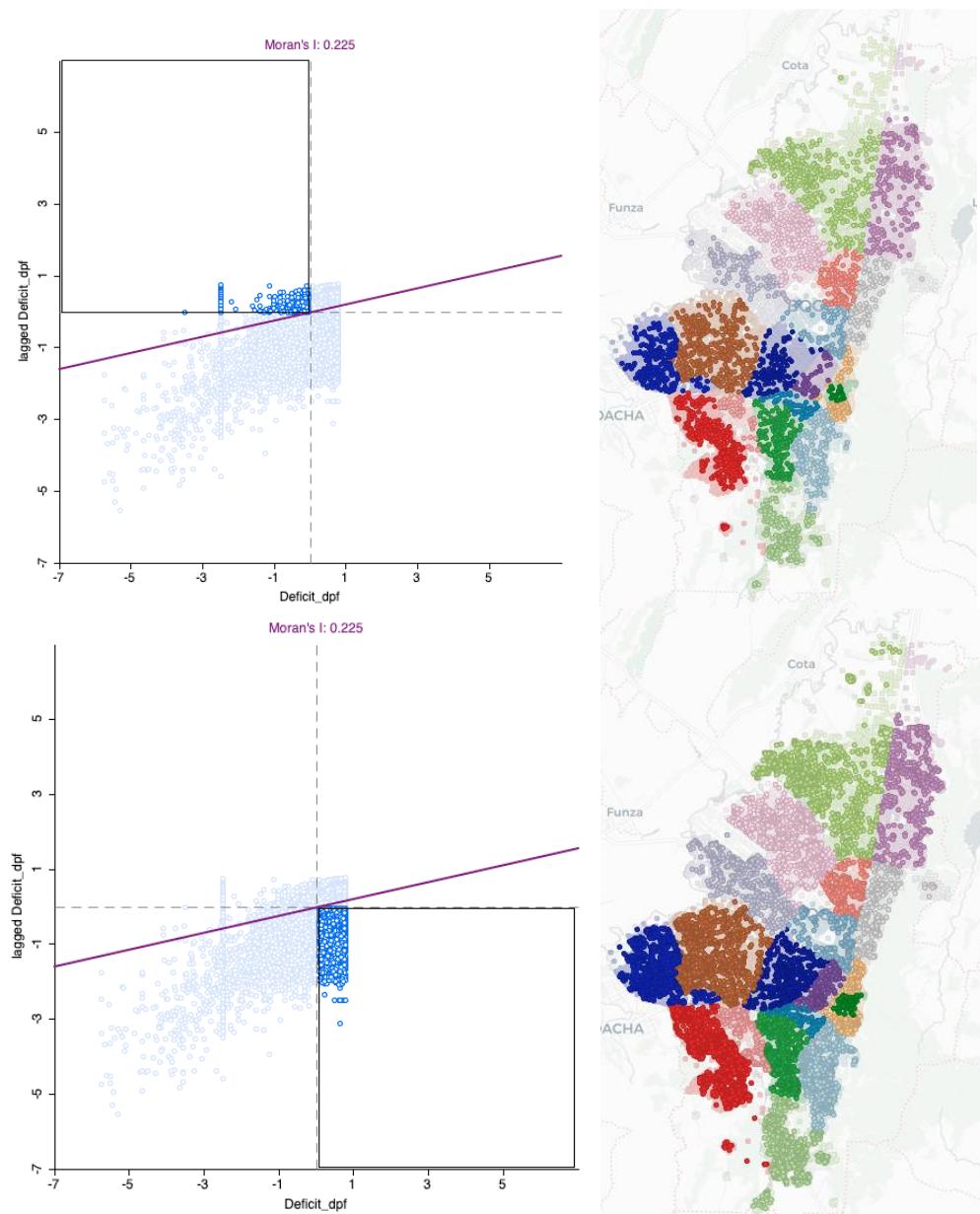
El gráfico de dispersión del índice de moran, permite visualizar las situaciones de los valores del indicador en el plano espacial, en la cual el primer cuadrante indican los valores que representan una mayor magnitud y en el tercero la menor magnitud del indicador, así mismo permite evaluar bajo la estadística deductiva los resultados cuyas interpretaciones solo se encuentran en una hipótesis nula, la cual establece que el indicador se distribuye de forma aleatoria entre las entidades del área de cada localidad, es decir que los procesos que promueven el patrón de valores se constituyen bajo una condición aleatoria. Debido a que el índice de morán se encuentra expresado en una magnitud de 0.225 aproximadamente 22.5%, indica que los datos en la totalidad presentan patrones completamente aleatorios en su distribución espacial comparada con la escala de -1 a +1, en la que el 1 positivo indica una alta auto-correlación espacial y el -1 expresa una negativa pero perfecta auto-correlación.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

Localidades con valores atípicos según el Local Moran (LISA):

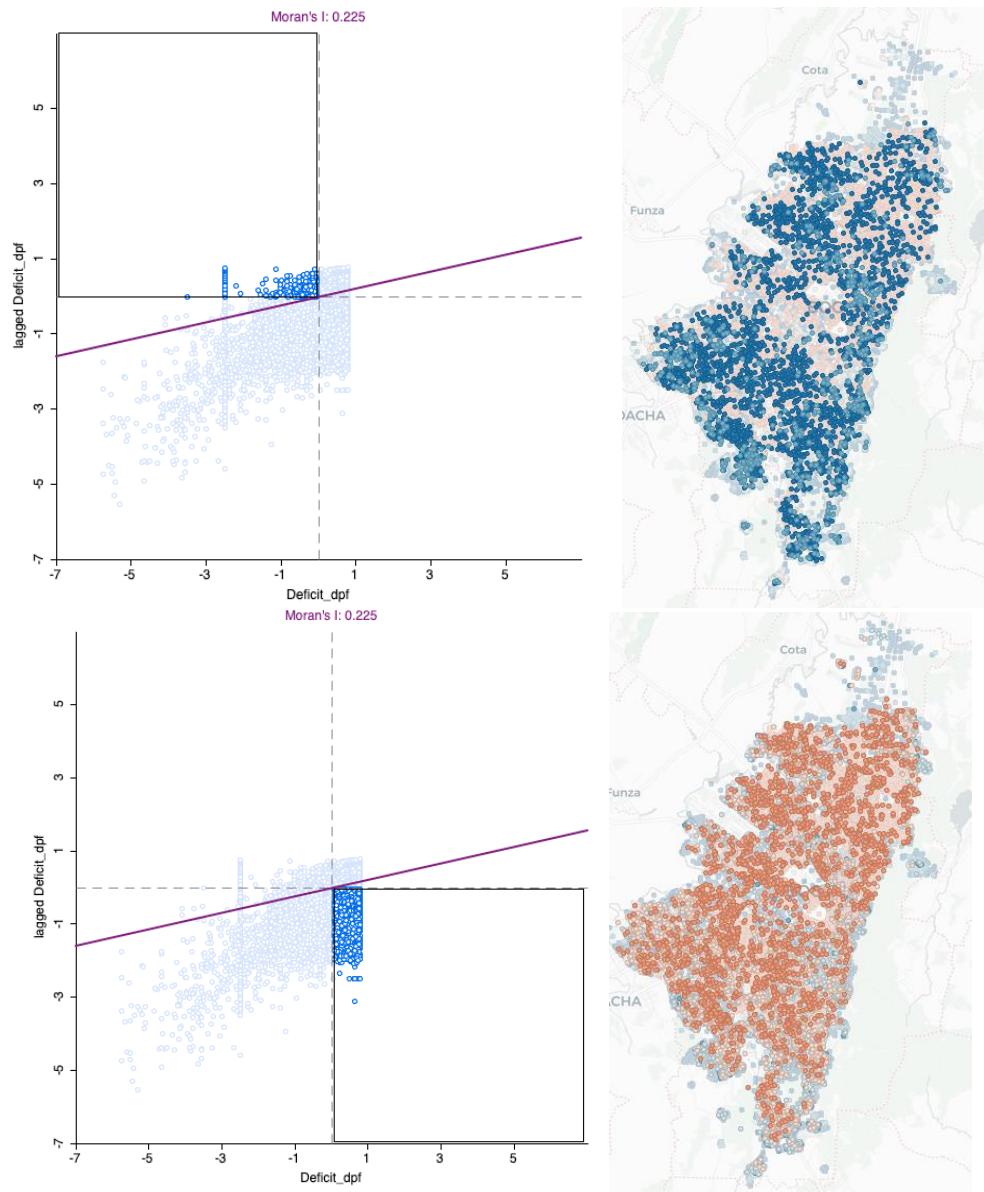
En el segundo y cuarto cuadrante, aquellos datos que se encuentren con condiciones anormales de distribución se presentan las localidades resaltadas en "highlights", con un grado de significancia del 95% y una cantidad de permutaciones de 999.

Ilustración 8. Mapas de contraste del DDIM con el gráfico único de localidades.



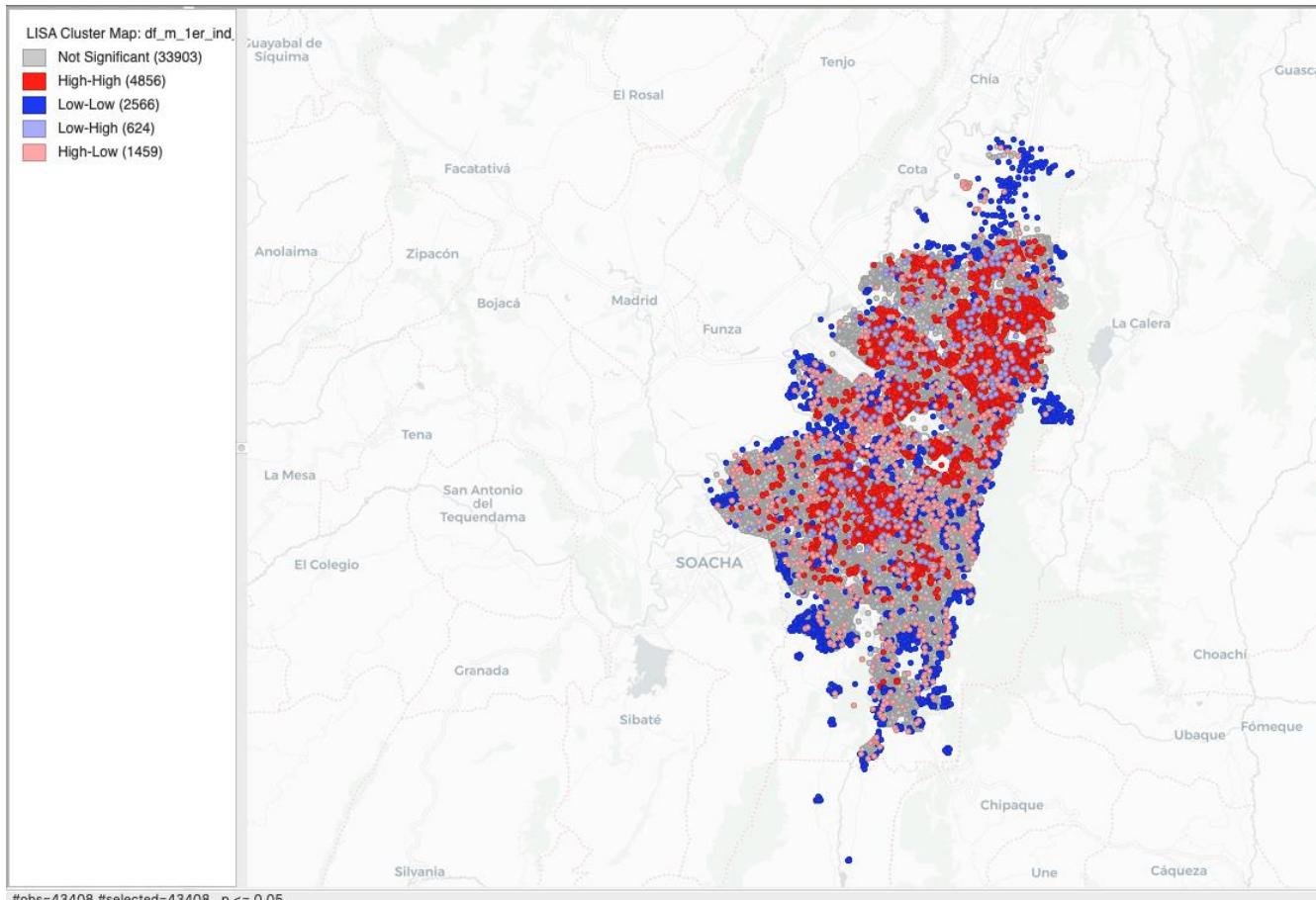
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

Ilustración 9. Mapas de contraste del DDIM con el gráfico de clasificación intercuartilica.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

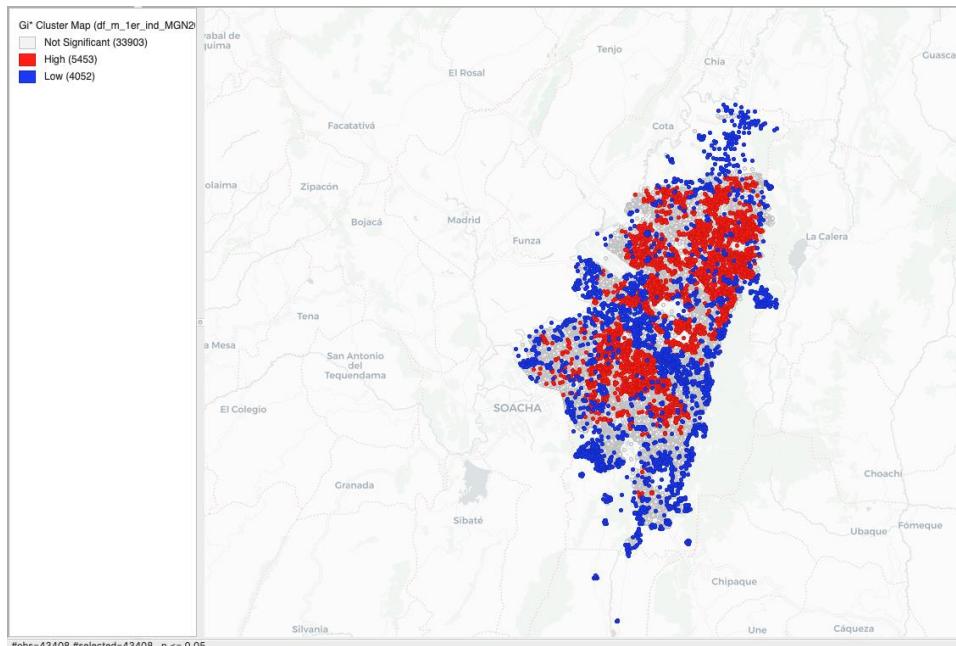
Los valores ubicados en el cuadrante II (Superior izquierdo), al ser contrastados con el gráfico que mide las proporciones en valores mayores al 1.5 veces el rango intercuartílico corresponden a los valores que se encuentran entre -1 y 0.585, y los valores en el IV (Inferior derecho) los que se encuentran entre 0.93 y 1, es decir Superávit.

Ilustración 10. Mapa de significancia y Clúster : Método de “Índice de Morán y LISA”.


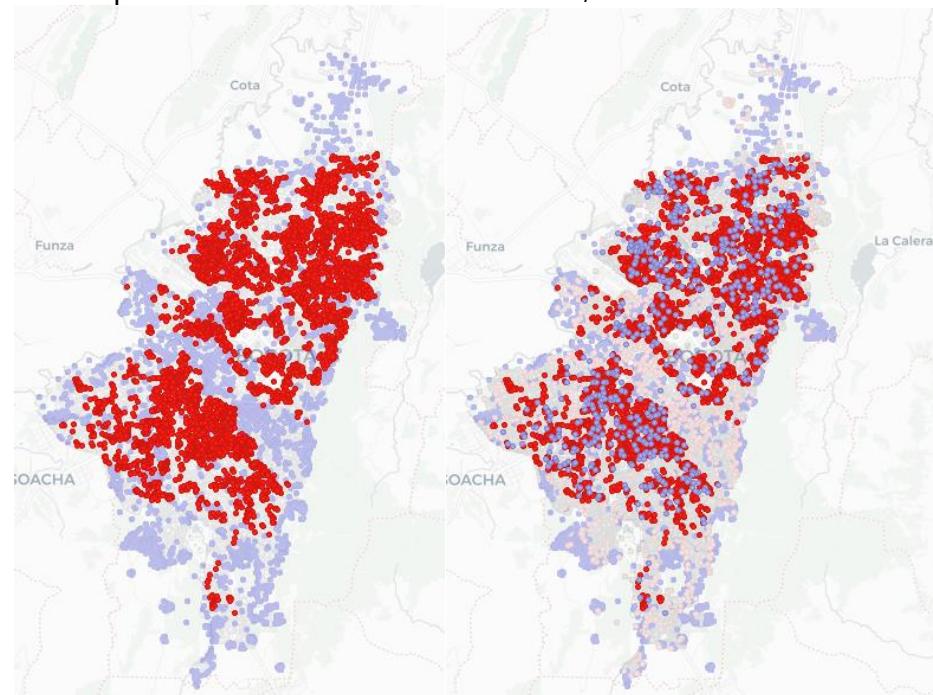
En la **Ilustración 10**, se observa que para las diferentes entidades de la capa que están formando algún tipo de agrupación espacial de tipo clúster y el nivel del valor para el indicador habitacional de entorno de significación estadística existen más de 600 manzanas que se agrupan en diferentes categorías significativas. Se conforman los clústeres donde para un menor valor del “P-value”, la probabilidad de encontrar manzanas con valor correlacionadas con la entidad de contigüidad próxima o los vecinos próximos y viceversa, o agrupaciones de manzanas que comparten características significativas y que ambas estén sujetas a condiciones significativas iguales de mediciones del indicador de valores altos se muestran con color “rojo” (Altos y altos) y las que tienen un grado de significancia mayor que el 0,05, se muestran de color azul en el mapa de clúster, esta misma de color gris en el mapa de significancia, con el 95% de confiabilidad contenido de 999 permutaciones de matriz de pesos de contigüidad, es decir que en el 95% de las ocasiones se presenta este patrón espacial siendo incorrecto el 5% de los casos.

Los resultados obtenidos anteriormente, se pueden contrastar con otras herramientas geo estadísticas para evaluar la pertinencia de los resultados del Modelo LISA. Para este caso se utiliza el método "Gi*", los resultados se explican a continuación.

Ilustración 11. Mapa de significancia y Clúster : Método de "estadístico Getis-ord Gi*".

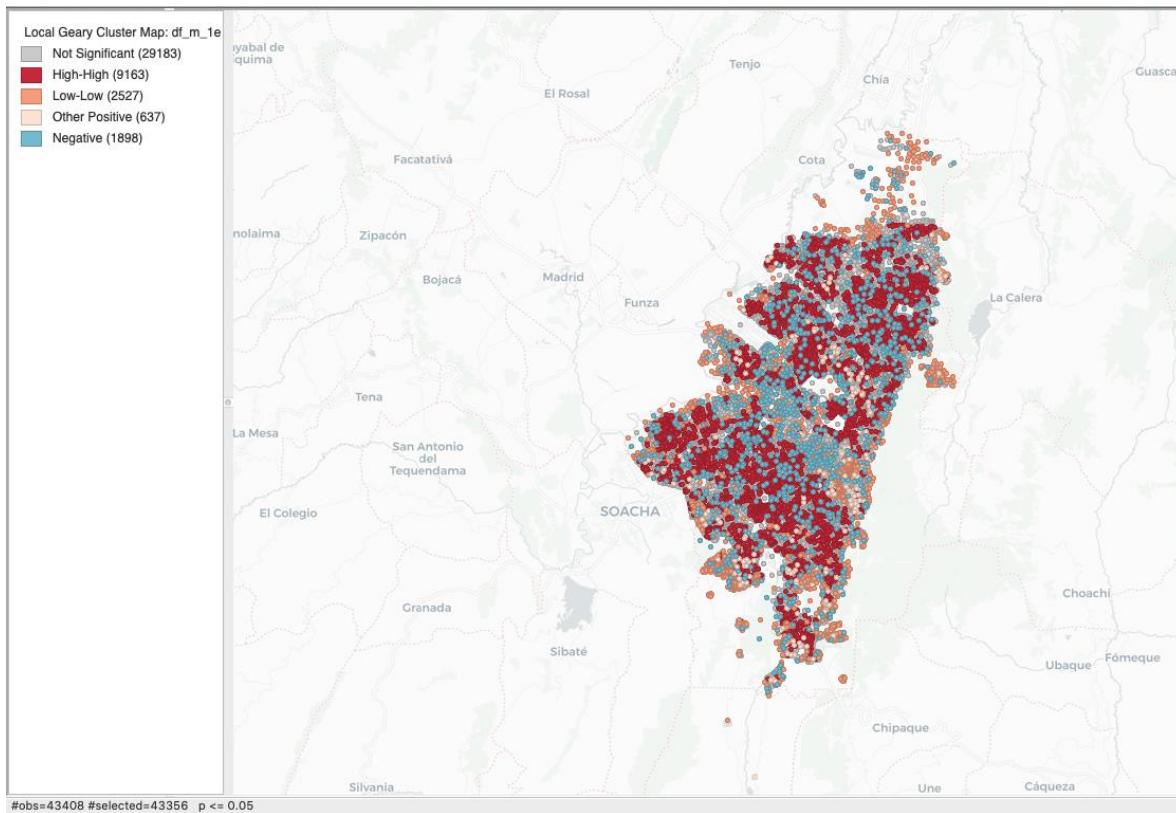


Comparando los valores altos de "Gi*", con el método "LISA":



En la visualización geográfica izquierda seguida de la **Ilustración 11**, se comparan dos métodos Geoestadístico representando los resultados obtenidos al seleccionar solo los valores con significancia alta del método Gi*, contrastando con mapa de la derecha los valores del método Índice de Morán (LISA) que fueron seleccionados, pero esta vez obteniendo una visualización para cuales de estos valores responden el método LISA, que en términos de agrupación se cumplen el criterio de alta significancia, no obstante, para el método LISA se encuentran los clústeres mejor discriminados de acuerdo a la concentración de valores con altos y bajos grados de significancia representados en las manzanas de colores morados.

Ilustración 12. Mapa de significancia y clúster : Método “Local Geary Clúster” .



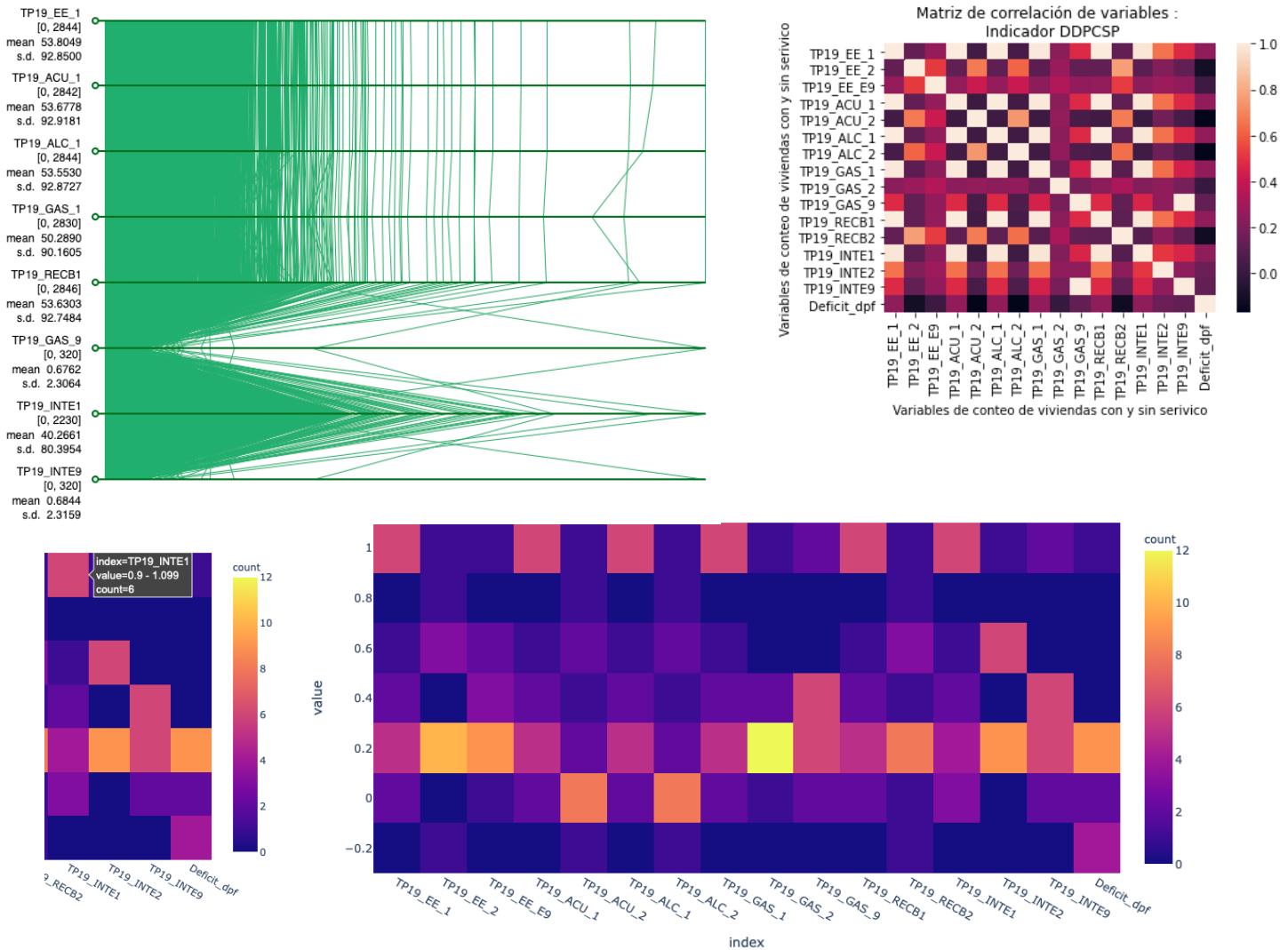
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

Con este método se obtiene que 9,163 manzanas están conformando el clúster con mayor significancia estadística, un total de 29,183 no significantes y 2527 con bajo grado de significancia.

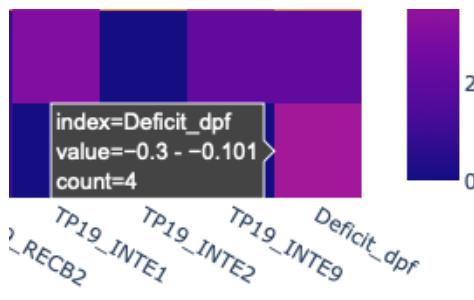
Comparando estos resultados conforme a las variables que permitieron construir el indicador a partir de la siguiente correlación, se explican los patrones para las entidades que demuestran mayor incidencia de auto-correlación con gráficos de densidad de correlación, matriz de correlación y paralelogramos: En el mapa de densidad de correlaciones, es posible observar que

de acuerdo a la matriz de correlaciones existen 8 (ocho) variables, las que en términos de escala de número de correlaciones altas observadas contienen de 2 a 6 variables correlacionadas con respecto a todas, pero que su valor de correlación esta en límites hacia el 1 (correlación alta positiva) estas son : "TP19_EE_1", "TP19_ACU_1", "TP19_ALC_1", "TP19_GAS_1", "TP19_REC_B1", "TP19_GAS_9", "TP19_INTE_1" y "TP19_INTE_9", por lo tanto se procede a analizar estas variables en un paralelogramo, en el cual se observan los flujos de valores y comportamiento que las variables presentan determinando similitudes a las variables que refieren de población favorecida, esto es que las condiciones estadísticas como lo son la media y la desviación estándar son parecidas.

Gráfico 9. Palalegrama, Matriz de correlaciones afines DDPCSP



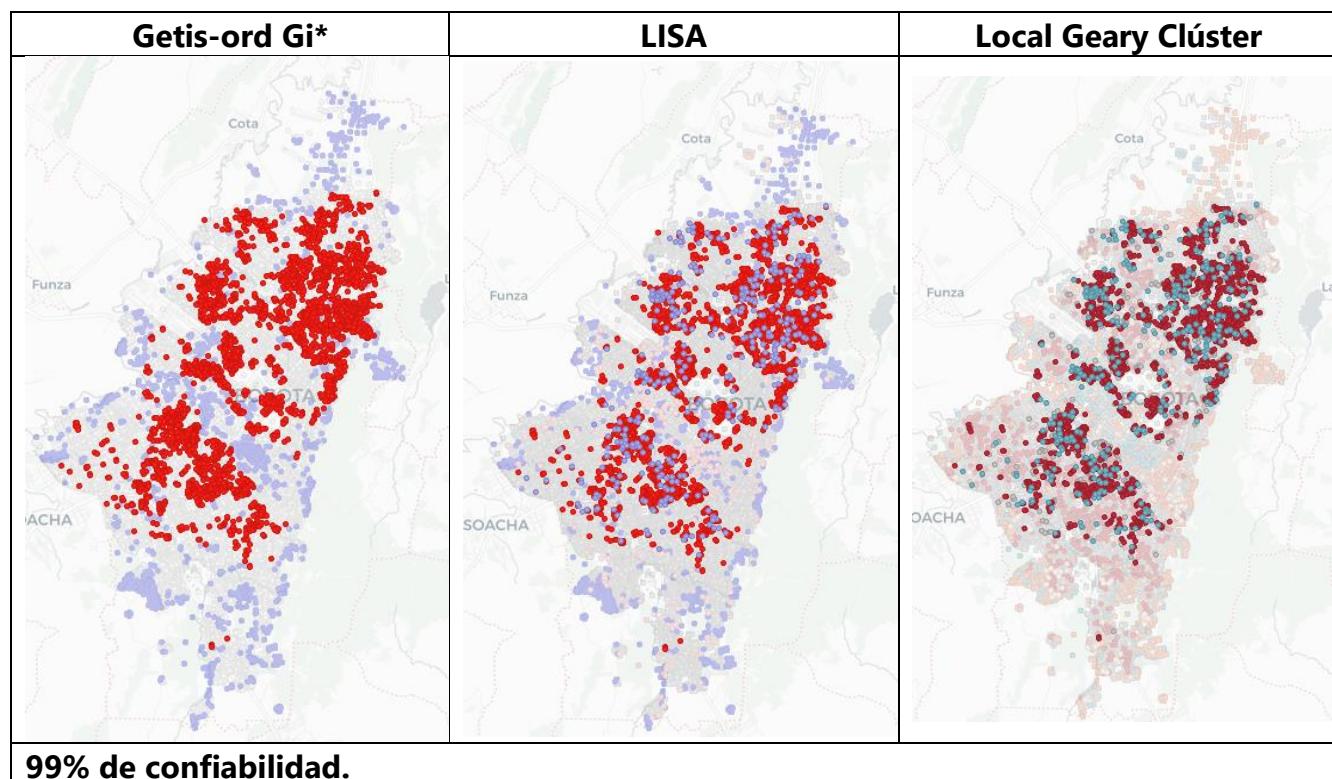
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA y Python, OIN DANE.



La variable "Deficit_dpf", correspondiente a el indicador DDPCSP, se correlaciona con 4 variables en los rangos -0.3 a -0.1, esto es una correlación débil y negativa, lo que es equivalente afirmar que en la medida que este indicador presente valores positivos las variables en cuestión (cuatro variables observadas) serían influenciadas en el sentido diferente por tanto estas disminuirían.

Al aumentar el grado de verosimilitud de manera que permita ver las manzanas con mayor significancia estadística en condiciones estudio de mayor confiabilidad, se seleccionan los valores con mayor grado de significancia en el método "Getis-ord Gi*" en comparativa con los otros métodos. Una vez aplicado los cambios, es posible observar la coincidencia de estos en el método "LISA", además de los valores que presentan clústeres "Bajo – Altos", y para el método "LGC" (Local Geary Clúster), se evidencia en la selección valores que según la escala de ese método manifiestan ser significativos.

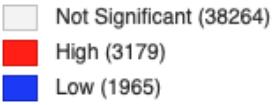
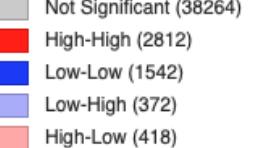
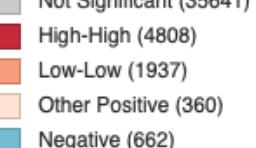
Tabla 12. Mapas de significancia del Indicador DDPCSP con grado de verosimilitud del 99%.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

Al aumentar el grado de confiabilidad es de notarse que para los tres métodos verificados, las manzanas que en términos de distribución geoespacial y con los valores del indicador en Superávit son las zonas con color "Rojo". Revisando este suceso se analizan las leyendas de los métodos de acuerdo a la cantidad de manzanas integradas para cada categoría:

Tabla 13. Leyendas de Mapas de significancia del DDPCSP.

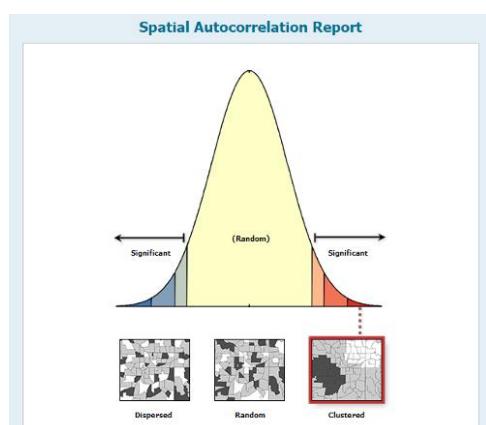
Getis-ord Gi*	LISA	Local Geary Cluster
		

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

A diferencia de los métodos de "Gi*" y "LISA", el método "LGC" tiene alrededor de 4808 manzanas que son de contigüidad "Alto-Alto" ("High-High"), siendo 1.7 veces mayor que la cantidad expresada por el método LISA, 1.51 veces mayor que el método Gi*.

El mismo fenómeno se estaría presentando al seleccionar los valores significativos de LGC,

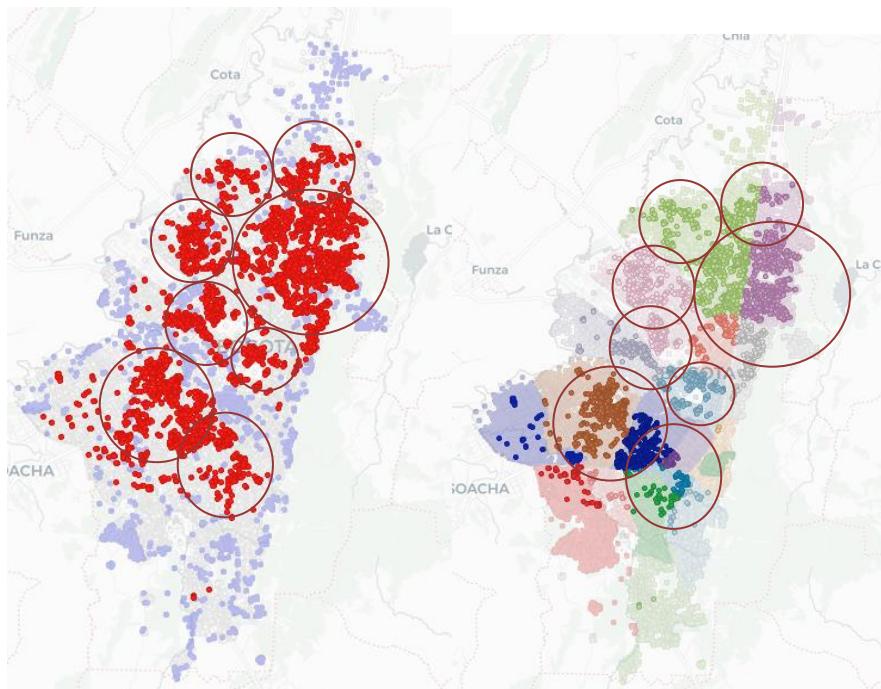
Ilustración 13. Representación de resultado de auto correlación espacial.



comparándose la coincidencia con los modelos del "Gi*" y "LISA" que resultan dar resultados muy próximos. Estos resultados de contraste son el efecto de poder inferir bajo los estadísticos que en el 99% de los casos es probable que esas agrupaciones no se deban a un factor del azar o de aleatoriedad sino que responden a un tipo de auto correlación espacial, las más comunes son:

+Para las manzanas en "rojo" representadas anteriormente hacen referencias a que como resultado de todas las permutaciones posibles, y luego de un proceso aleatorio, inminentemente estas zonas tengan este patrón de agrupación. En ese orden de ideas, se hace referencia a que conforme a las condiciones de aleatoriedad de la distribución de los datos del indicador cuando se muestran puntos azules "ColdSpots", estos hacen referencias a las observaciones con

perfíles similares en términos estadísticos, separadas gradualmente y además que no representan un peso estadísticamente significativo que permita agruparse entre ellos.

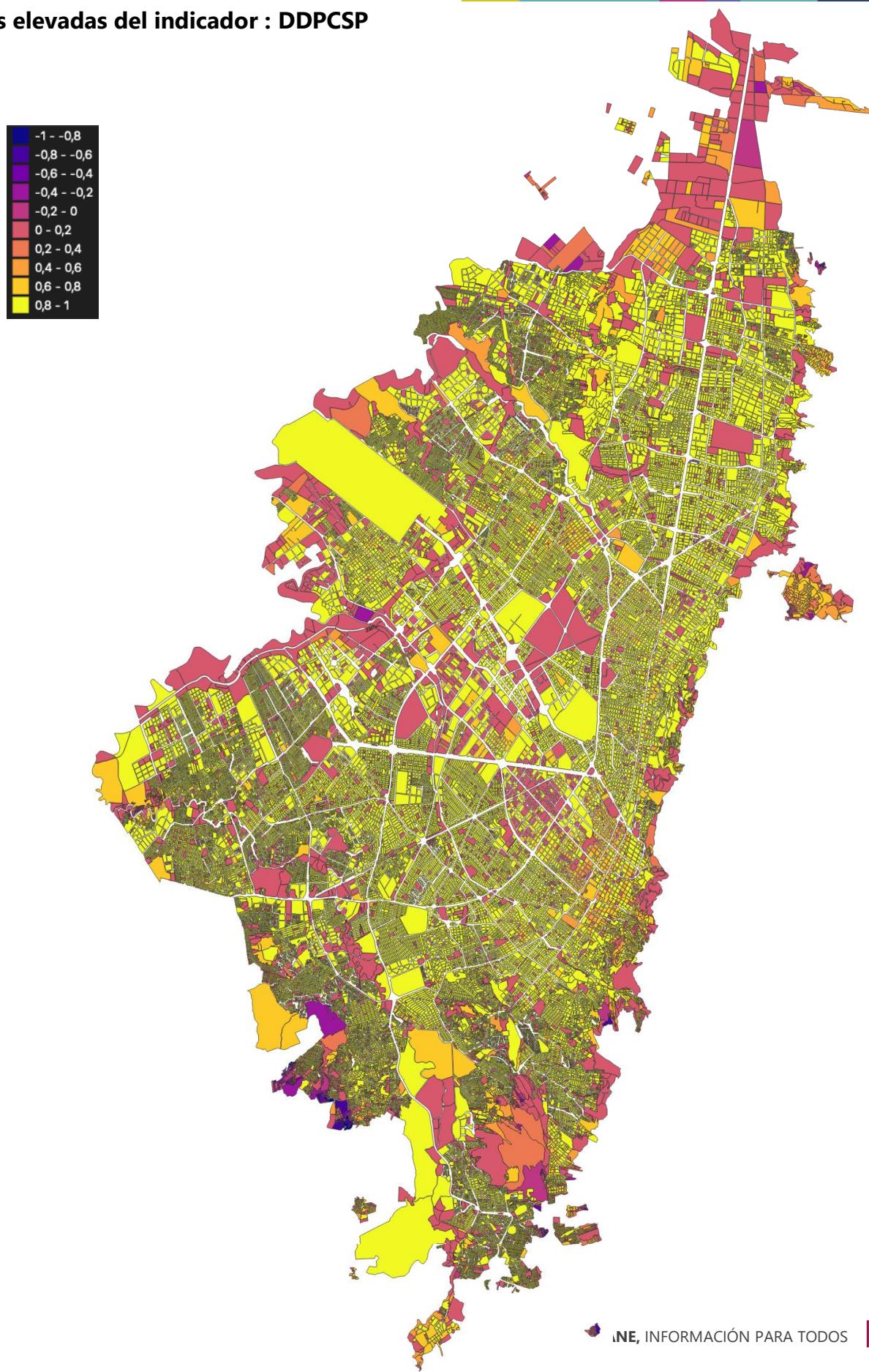
Ilustración 14. Mapas de significancia estadística método Gi* vs Mapa único de localidades.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

En el 99% de los casos existe mayor población con cobertura en servicios públicos que las que no, son en las localidades de "Suba", "Usaquén", "Kennedy", "Engativá", "Puente Aranda" y "Teusaquillo", esto refiere a que de acuerdo a los valores estáticos del marco Geoestadístico nacional en estas localidades mencionadas para los vectores de datos visualizados en los gráficos geoespaciales de **Ilustración 14**, se infiere la existencia bajo una alta confiabilidad de la posibilidad de encontrar manzanas con "Alto superávit" rodeadas de manzanas con alto "Superávit".

Ilustración 15. Análisis de las manzanas con mediciones elevadas del indicador : DDPCSP

NOMBRE DE LA INVESTIGACIÓN / TEMÁTICA

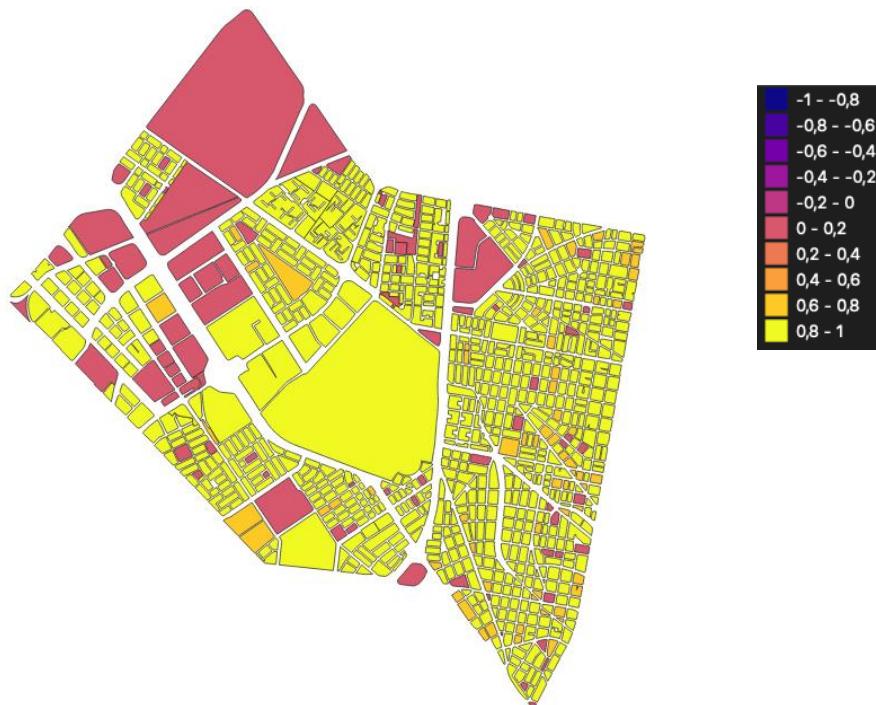


En la clasificación del indicador en la ciudad de Bogotá DC por manzana, es de notarse la distribución de las manzanas conforme a sus colores dados por la clasificación una alta presencia de manzanas con colores para valores entre 0.8 y 1, esto quiere decir que existen muchos registros de manzanas con altas mediciones que se encuentra en "Alto Superávit" y "Crítico Superávit", la población favorecida estas entre 80% y el 100% en cada manzana.

Análisis de la localidad de Teusaquillo con mediciones del indicador :

En la localidad de Teusaquillo, se presentan manzanas con valores dentro de las escalas del [0-1], cuya presencia de datos se encuentran en la categoría de "Equilibrio" [0,0.2] ubicadas al Norte de la localidad y en el extremo "Noroeste" de la ciudad, cuyos polígonos son los más grandes.

Ilustración 16. Mapa localidad Teusaquillo con distribución geoespacial del indicador DDPCSP.



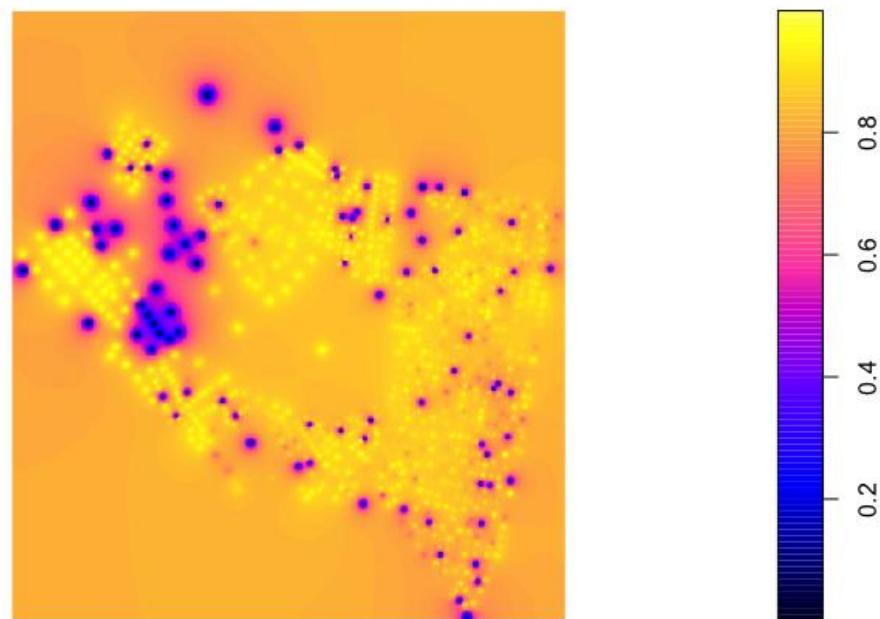
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE.

En el barrio de "La Esmeralda", la mayoría de manzanas se ubican en el rango de 0.8 a 1 valores del indicador, esto es que existen manzanas que se encuentran en la categoría "Moderado Superávit" tal como lo son la manzana central con color naranja cuya medición del índice es equivalente a "0.6956", pero también se encuentra la manzana con medición equivalente a 0 (cero) clasificada dentro de la categoría "Equilibrado", la población esta dividida ya que la mitad cuenta con servicios y la otra mitad no cuenta con servicios en esta categoría.

Ilustración 17. Mapas de selección de manzanas en el barrio “La Esmeralda”

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE.

En la interpolación con kriging calculada con la covarianza de las magnitudes del indicador, se contemplan los valores de manzanas de la localidad de Teusaquillo en Bogotá DC, es posible realizar un comparativo entre las zonas de esta población considerando un escala de máximos y mínimos de acuerdo a la medición del indicador en esta segmentación geográfica.

Ilustración 18. Mapa de interpolación kriging del indicador DDPCSP.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Se contemplan un gran cantidad de vectores en los que los indicadores presentan altas magnitudes comparados con las manzanas en la misma localidad, estos valores por encima del promedio podrían ser considerados en "Superávit".

Análisis prospectivo de la localidad de Teusaquillo con mediciones del indicador : Auto-correlación espacial.

El método KNN (K-nearest neighbors) es empleado para construir la matriz de pesos de contigüidad en la base de datos filtrada para la localidad de Teusaquillo, cuyo método de cálculo de proximidad empleado es "Euclíadiano", las cantidades de vecinos son de 4 por defecto y con un total de 966 manzanas presentadas en esa localidad.

Bajo un 95% de confiabilidad es posible afirmar que en el Barrio de "Ciudad Salitre", se concentran las manzanas que representan significancia estadística, es decir que se acepta la idea que para las manzanas que están por debajo del p-valor igual a 0.05, son manzanas que responden a algún tipo de agrupación espacial debido al peso de contigüidad de las manzanas vecinas, a menor valor de "p" mayor será la aceptación que estas agrupaciones no se deben a factor aleatorio, sino por el contrario por influencias de las naturaleza del predio.

Ilustración 19. Mapa de significancia estadística método LISA : Localidad Teusaquillo.

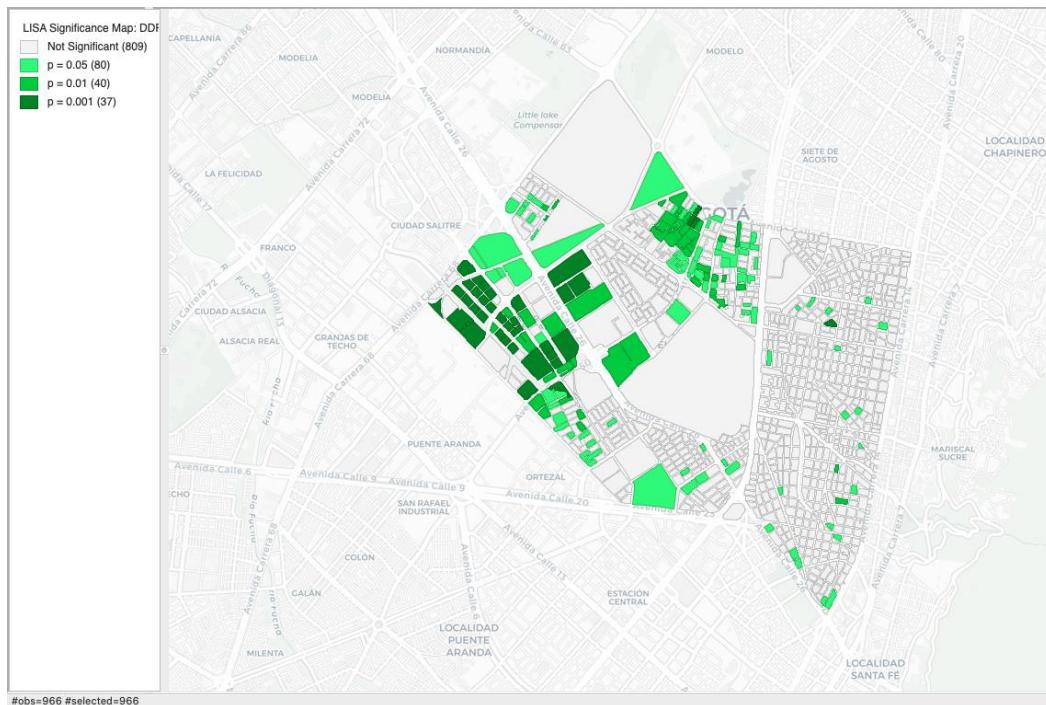
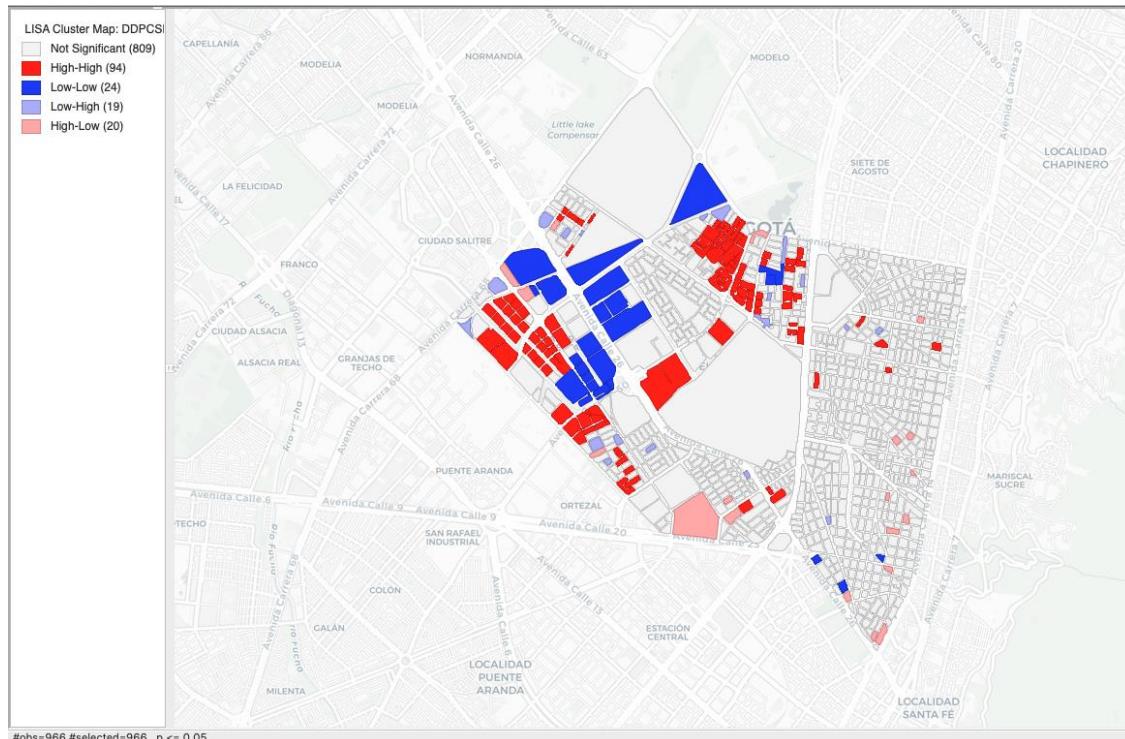


Tabla 14. Resumen del método de creación matriz de pesos de contigüidad.

Property	Value
type	k-NN
symmetry	asymmetric
file	DDPCSP_TEUSAQUILLO.gwt
id variable	field_1
distance metric	Euclidean
distance vars	centroids
neighbors	4
# observations	966
min neighbors	4
max neighbors	4
mean neighbors	4.00
median neighbors	4.00
% non-zero	0.41%

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de GeoDA, Observatorio Inmobiliario Nacional DANE. En la **Ilustración 20**, se presentan 94 manzanas con significancia estadística, estas distribuidas en diferentes barrios señalando la fuerte existencia de la posibilidad de encontrar efectivos del indicador en distintas zonas de la localidad, que contengan un superávit del índice de servicios o que la población con cobertura a servicios públicos sea proporcionalmente mayor a las que no, en el barrio "Ciudad Salitre Oriental" y "Pablo VI".

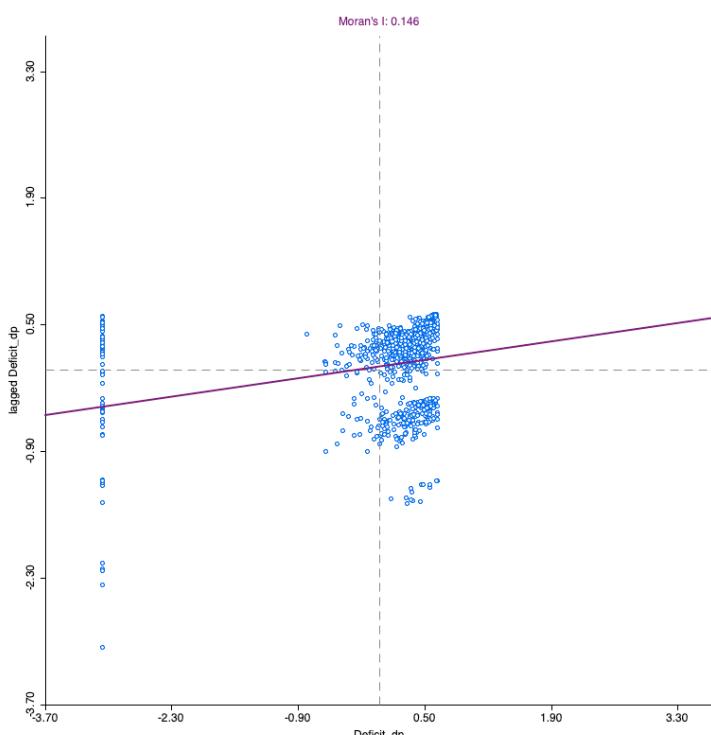
Ilustración 20. Mapa de clúster método LISA : Localidad Teusaquillo.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de GeoDA, Observatorio Inmobiliario Nacional DANE.

El método Local Morán funciona en excelente condiciones para soportar la premisa que en el 95% de los casos las manzanas agrupadas en la población con cobertura a servicios públicos predomina en las manzanas de color rojo, y en las manzanas de color azul hay mayor agrupación de manzanas con concentración poblacional con valores bajo del indicador comparado en la misma localidad, en este caso las manzanas que se encuentran en "Equilibrio", debido a la escala que se presenta en la clasificación de los valores obtenida de forma preliminar.

Gráfico 10. Gráficos de dispersión Índice de Morán.



El índice de morán se encuentra expresado bajo una magnitud de 0,146 aproximadamente 14,6%, esto indica que los datos en su totalidad presentan patrones positivos, es decir existe una precaria auto-correlación espacial, haciendo referencia a una baja probabilidad encontrar agrupaciones dentro de la distribución espacial de los datos en los que coincidan las mediciones y sean considerados valores representativos de las zonas sean "hotspots" (Valores de manzanas en Superávit rodeadas de manzanas con Superávit), o manzanas en "ColdSpots" (para este caso en Teusaquillo, valores de manzanas en Alto Superávit o en Equilibrio acompañados o rodeados de manzanas con Superávit o Equilibrio).

Hallazgos:

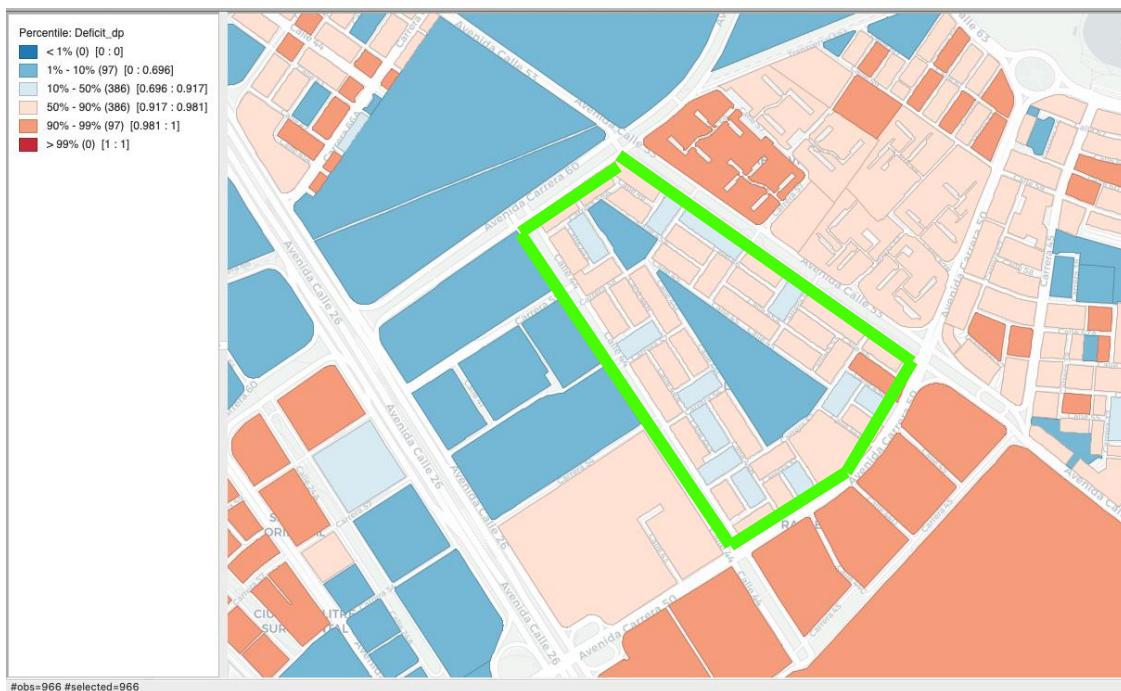
- 1)** De los métodos empleados las entidades que coinciden en los niveles de significancia estadística responden a una alta probabilidad que en un proceso aleatorio de los valores del indicador se armen grupos en esas zonas con igualdad de características, pero del resto resultan triviales en significancia o siguiendo una aleatoriedad.
- 2)** Los valores "No significancia" cuyo nivel aparece en la leyenda, enmarca las manzanas que se ajustan en la distribución normal teórica pero con un grado de contigüidad nulo para una posible agrupación, por lo tanto existe una condición bastante aleatoria de la distribución geoespacial.
- 3)** El barrio de "La Esmeralda" presenta una distribución de desviación estándar muy informe, a excepción de los polígonos con niveles bajos del indicador o en categoría "Equilibrio".

Ilustración 21. Mapas de desviación estándar del índice DDPCSP en el Barrio : "La Esmeralda".



4) En el mapa geoespacial de percentiles **21b**, la clasificación de las manzanas en el barrio, muestra los niveles de poblaciones favorecidas en el barrio "La Esmeralda", presentando 12 manzanas con valores entre 0.696 y 0.917, es decir "Medio Superávit", "Moderado Superávit" y "Alto Superávit", distribuidas de forma aleatoria sin ningún tipo de clúster.

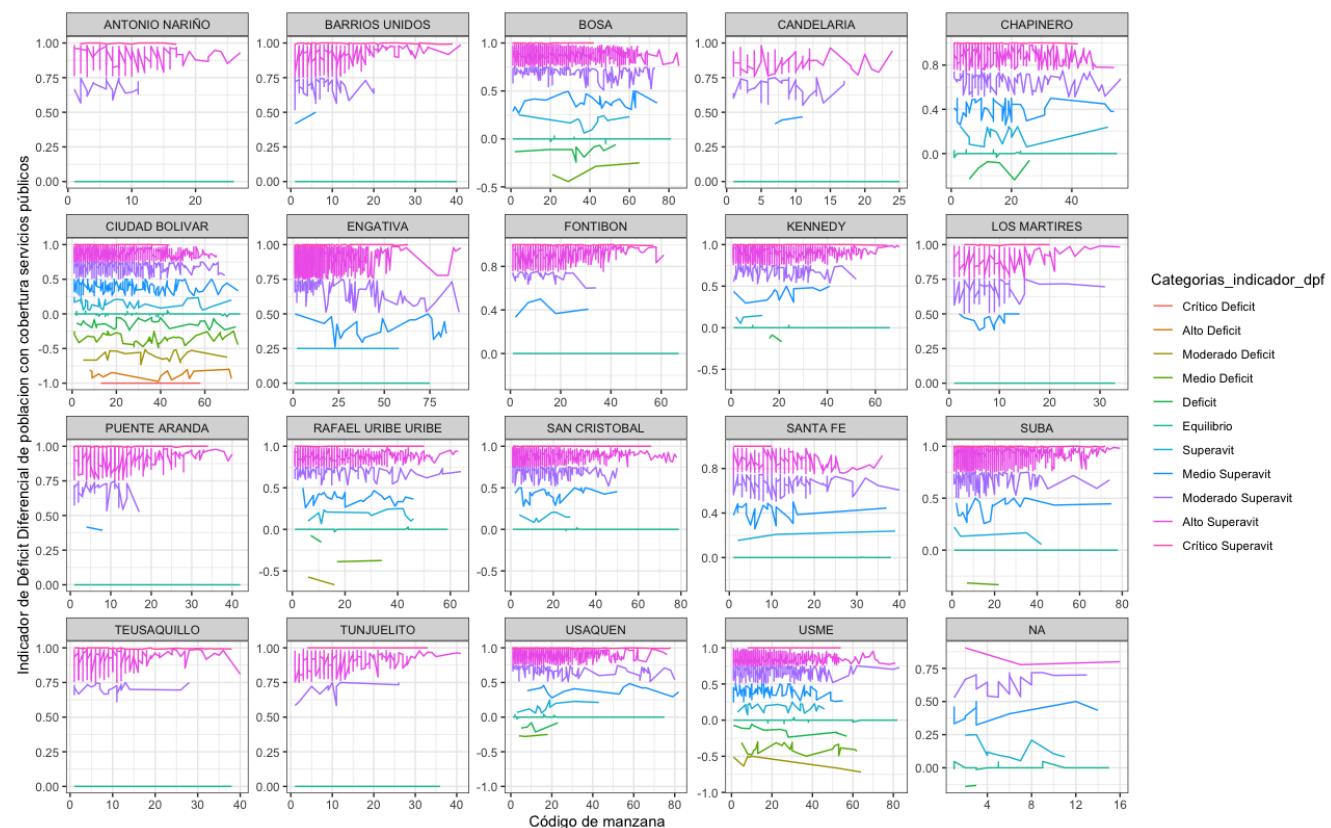
Ilustración 21b. Mapas de percentiles del índice DDPCSP en el Barrio : "La Esmeralda".



(Minitab Blog Editor, 2016)

Los valores del indicador para cada localidad se representan en la **Tabla 15**. El comportamiento del indicador en la localidad de Teusaquillo se encuentra en un aspecto estacional, con una amplitud elevada en función de la manzana obteniendo un total de 40 manzanas enumeradas y repartidas en todo el data set. Las manzanas con el código del 1 al 20 muestran un patrón estacional ascendente, en la que los registros del indicador están por encima del 0.75 y menores que 1, además se encuentra que los datos entre la manzana 20 y 40 presentan un patrón estacional descendente pero menos uniforme debido a que los tamaños de estas manzanas varían y se encuentran distribuidas bajo una gran dispersión. En general, los promedios son diferentes para todas las localidades, por ejemplo es de apreciarse que en la localidad de "Ciudad Bolívar" los registros por debajo de -0.75 en el indicador tienen presencia desde la manzana 20 a la manzana 60.

Tabla 15. Gráfico de series para las manzanas en cada localidad : Niveles del DDPCSP en términos de categorías.

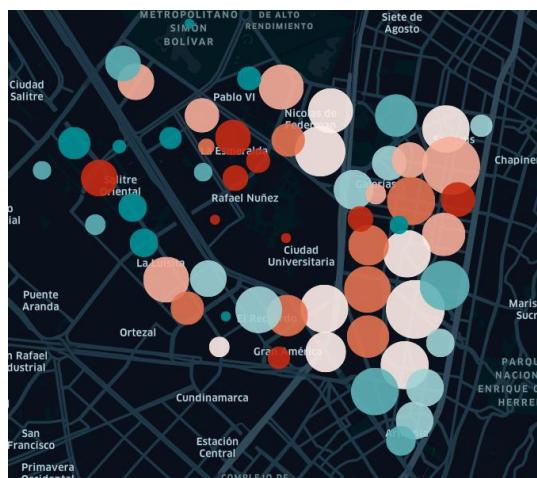


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

La extensión a visualizaciones que permitan analizar los por mayores y la distribución del indicador resultan útiles a la hora de tomar decisiones, en este caso se muestra un gran interés

por profundizar los métodos de agrupación que permitan representar los datos, por tanto a continuación se presenta un análisis de clúster.

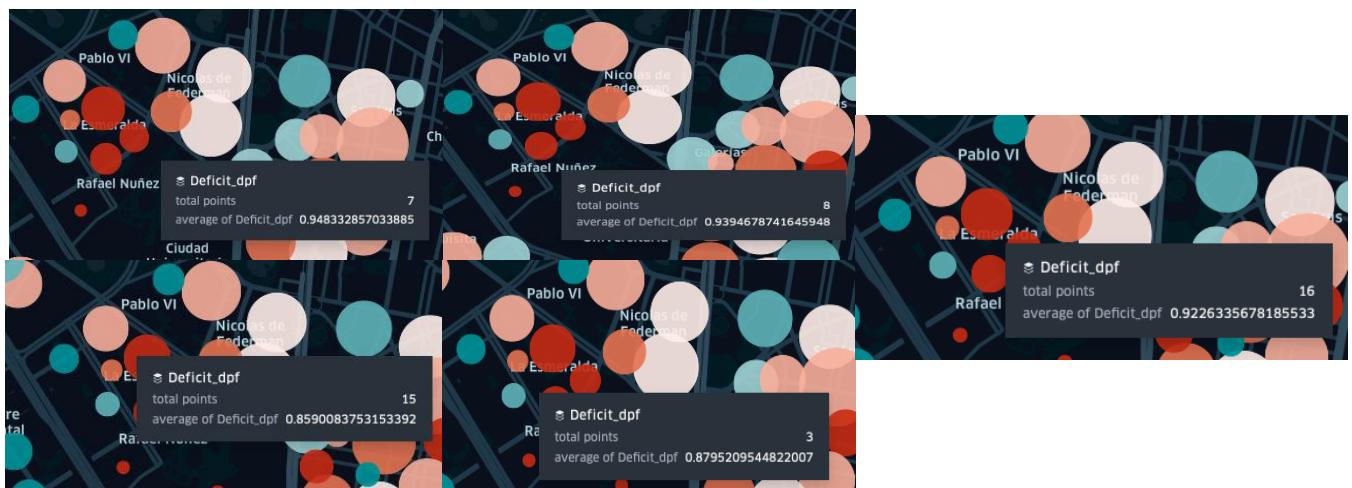
Ilustración 22. Mapa de clúster localidad teusaquillo (Operación promedios entre manzanas).



con cobertura en servicios públicos.

+ A un radio de 35 metros de distancia a la redonda para cada registro de manzana se obtiene un mapa que representa los clústeres, para el barrio de "La Esmeralda" procesando los datos se obtiene un total de 5 clústeres; una cantidad equivalente de 3 clústeres de tipo "Alto Superávit", estos tienen diferentes cantidades de manzanas agrupadas, es de apreciarse en la **Ilustración 22** los puntos de color "Rojo". Un primer clúster de 8 manzanas en promedio del nivel de índice para tal clúster es de 0.9483, es decir que más del 90% de la población en esa manzana poseen cobertura a servicios públicos, y 2 manzanas que una vez calculado el promedio se encuentran entre 0.85 y 0.87, es decir tienen mayor presencia de poblaciones en promedio que no cuentan

Ilustración 23. Mapas de clústeres de 5 clústeres.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde Python, Observatorio Inmobiliario Nacional DANE.

2.2 Construcción del indicador: “Déficit en cobertura de servicios públicos” (DCSP).

El déficit en cobertura de servicios públicos, refiere de la existencia de estados de carencia en la satisfacción de las necesidades de la población conforme al acceso a servicios públicos, cuya medición busca caracterizar que zonas poblacionales cuentan con la mayor cantidad de servicios públicos y cuales no cuentan en su totalidad con conexión a ningún tipo de servicios.

Desarrollo del indicador: DCSP

En este apartado se presenta el proceso empleado para la creación del indicador del Déficit en cobertura de servicios públicos.

Seleccionando las columnas para el indicador, se emplea la siguiente metodología:

Tabla 16. Resumen estadístico de variables afines al DCSP

— Data Summary —————	
	Values
Name	data
Number of rows	43408
Number of columns	7
<hr/>	
Column type frequency:	
numeric	7
<hr/>	
Group variables	None
<hr/>	
— Variable type: numeric —————	
skim_variable	n_missing complete_rate mean sd p0 p25 p50 p75 p100 hist
1 TP19_EE_1	0 1 53.8 92.9 0 16 35 60 2844 ━━
2 TP19_ACU_1	0 1 53.7 92.9 0 16 34 60 2842 ━━
3 TP19_ALC_1	0 1 53.6 92.9 0 15 34 60 2844 ━━
4 TP19_GAS_1	0 1 50.3 90.2 0 14 32 56 2830 ━━
5 TP19_RECBI	0 1 53.6 92.7 0 16 34 60 2846 ━━
6 TP19_INTE1	0 1 40.3 80.4 0 9 23 42 2230 ━━
7 TVIVIENDA	0 1 58.0 104. 0 17 37 63 4618 ━━

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

En este indicador se busca medir la incidencia de la composición total sobre la cantidad de servicios que la población tiene acceso para un plano estático de medida temporal. Es de apreciarse que en promedio las poblaciones de variables que según el diccionario de datos indican que son poblaciones que cuentan con servicios, están por debajo al promedio de conteo de viviendas totales por manzanas, esto debido a que los resultados de estas mediciones resultan estar clasificadas en 3 categorías, la población favorecida, población no favorecida y la población sin información relativa.

Para conocer que poblaciones cuentan con un nivel completo de cobertura de servicios, se establece la siguiente operación:

1). Establecer la participación por cada tipo de servicio en función de la cantidad total de viviendas por manzanas. Para esto es necesario considerar la variable "TVIVIENDA" que hace referencia a las cantidades de viviendas totales por manzana:

Ecuación 13. Consideración de la variable de conteo de vivienda "TVIVIENDA".

$$\text{Población total por manzana PTM} = \text{TVIVIENDA}$$

2). Se realiza la sumatoria de las proporciones calculadas a partir de cada uno de los servicios (variables afines).

Ecuación 14. Operación para la construcción del indicador DCSP (Calculo del SPS: General)

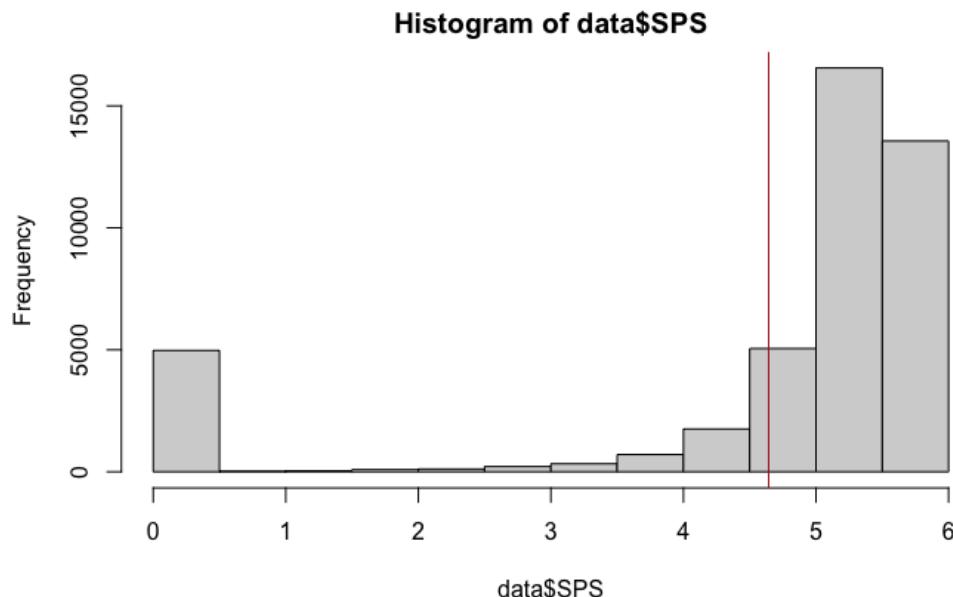
$$\text{Sumatoria proporciones por servicios SPS} = \sum \frac{\text{Conteo de viviendas por tipo "x" servicio}}{\text{PTM}}$$

Esta anterior **Ecuación 14** indica que cada proporción, es decir cada fracción de conteo de viviendas por cada tipo de servicio ("x") sobre la población total por manzana serían atadas a una operación de sumatoria, a nivel de detalle es posible apreciar a nivel de detalle la operación, respectivamente en la **Ecuación 15**.

Ecuación 15. Calculo del SPS :Detallado

$$SPS = \frac{TP19_EE_1}{PTM} + \frac{TP19_ACU_1}{PTM} + \frac{TP19_ALC_1}{PTM} + \frac{TP19_GAS_1}{PTM} + \frac{TP19_RECB1}{PTM} + \frac{TP19_INTE}{PTM}$$

Al efectuar estas operaciones se obtiene el siguiente histograma de representación de datos del SPS:

Gráfico 11. Histograma de datos SPS.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

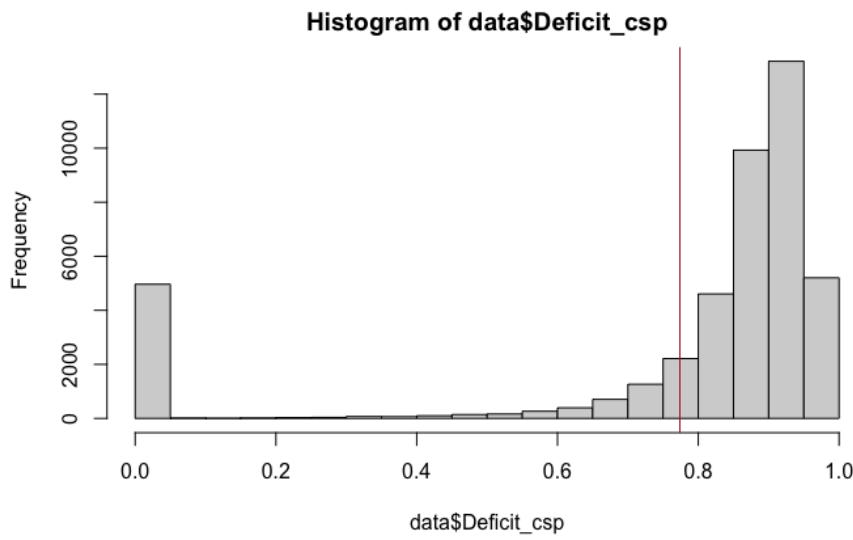
El **Gráfico 11** línea roja indica el valor de la media correspondiente del SPS es 4.6, es posible inferir que en toda la ciudad de Bogotá DC, en promedio por cada manzana tienen más de 4 servicios.

3). La ultima operación, es considerar una escala única que permita representar estos valores desde 0 al 1, el cuál indica que para un valor de indicador igual a 1, es porque en esa manzana se cuenta con todos los servicios y para un valor de cero, la manzana no cuenta con ningún tipo de servicio.

Ecuación 16. Cálculo del indicador DCSP.

$$\text{Deficit cobertura en servicios públicos por manzana (DCSP)} = \frac{SPS}{6}$$

Una vez elaborado el indicador se procede entonces a su análisis. A continuación se presenta el resumen estadístico y su histograma:

Gráfico 12. Histograma del indicador DCSP

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.8071	0.8871	0.7738	0.9274	1.0000

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

La distribución de los datos en el histograma del **Gráfico 12** muestran un asimetría, en el que la mayor cantidad de datos se encuentra en los rangos entre 0.8 y 1, es decir aquellas poblaciones que cuenten con más de 4 servicios.

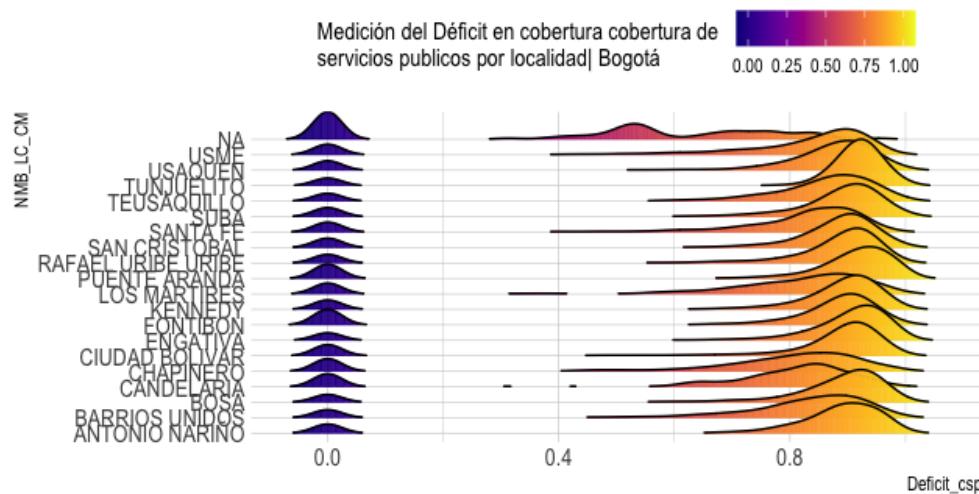
Categorización del indicador "Déficit en cobertura de servicios públicos": Análisis de datos.

Empleando los algoritmos afines para la construcción del análisis del indicador con tablas cruzadas, se emplea la categorización a partir de la escala de medida que permita medir de forma detallada la incidencia del indicador en cada manzana, para el cual se establecen límites y clasificación de la medición del mismo. Oscilando las métricas en los siguientes rangos:

- [0, 0.2] → "Alto Déficit en cobertura de servicios públicos"
- [0.2, 0.4] → "Medio Déficit en cobertura de servicios públicos"
- [0.4, 0.49] → "Déficit en cobertura de servicios públicos"
- [0.49, 0.51] → "Equilibrio en cobertura de servicios públicos"
- [0.51, 0.60] → "Superávit en cobertura de servicios públicos"
- [0.6, 0.8] → "Medio Superávit en cobertura de servicios públicos"
- [0.8, 1] → "Alto Superávit en cobertura de servicios públicos"

Estas categorías creadas con la finalidad de discriminar en mejor medida los valores de indicador.

Gráfico 13. Gráfico de densidad del indicador DCSP por localidad en la ciudad de Bogotá DC.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Todas las localidades en términos de la densidad presentan comportamiento de los datos similares, a excepción de las localidades que no se encuentran registradas, así mismo es de apreciarse que por ejemplo para localidad de "Ciudad Bolívar", tiene una curva de densidad de los valores concentrados próximos a 0.8 de tipo platicúrtica, esto refiere que los datos tienen mayor dispersión y por tanto es posible encontrar una significativa proporción de datos ubicados entre los rangos 0.4 y 0.9.

Ilustración 24. Resumen de cantidad de manzanas ubicadas en cada categoría DCSP

Alto Deficit	Medio Deficit	Deficit	Equilibrio	Superavit	Medio Superavit	Alto Superavit
5005	202	188	68	415	4563	32967

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Resumiendo las cantidades de manzanas por categoría, se calcula que el 75% de los manzanas se encuentran en la categoría de tipo "Alto Superávit", es decir que la mayor concentración de manzanas indican tener entre 4 y 6 servicios.

Análisis a partir de la tabla de contingencia del indicador DCSP

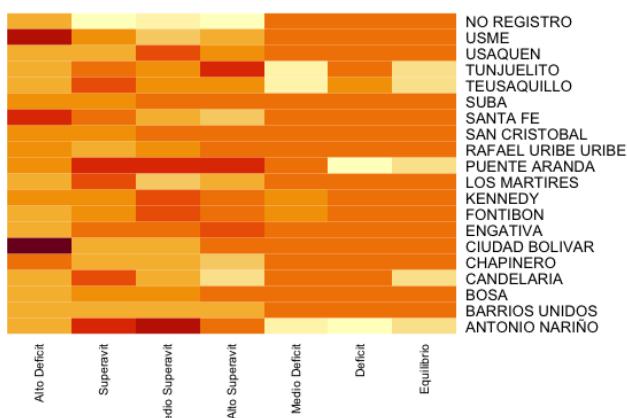
Teniendo en cuenta el promedio de indicador por localidad, la tabla cruzada calculada es la presentada en la **Tabla 17**.

Tabla 17. Tabla de contingencia del indicador para cada localidad.

NMB_LC_CM <chr>	Alto Deficit <dbl>	Superavit <dbl>	Medio Superavit <dbl>	Alto Superavit <dbl>	Medio Deficit <dbl>	Deficit <dbl>	Equilibrio <dbl>
ANTONIO NARIÑO	0.000000e+00	0.5777778	0.7583616	0.9038115	0.0000000	0.0000000	0.0000000
BARRIOS UNIDOS	0.000000e+00	0.5604106	0.7326241	0.8846310	0.3174603	0.4453214	0.4993221
BOSA	6.699077e-05	0.5619227	0.7383846	0.9077348	0.3364684	0.4472934	0.5011111
CANDELARIA	0.000000e+00	0.5714286	0.7318737	0.8709509	0.3111111	0.4244444	0.0000000
CHAPINERO	8.028259e-04	0.5584769	0.7287750	0.8779353	0.3602329	0.4566318	0.5000627
CIUDAD BOLIVAR	4.449505e-03	0.5558288	0.7322160	0.9063836	0.3153955	0.4562635	0.4998926
ENGATIVA	0.000000e+00	0.5686767	0.7428374	0.9169610	0.3534348	0.4529927	0.5000000
FONTIBON	1.904261e-04	0.5649841	0.7467473	0.8996498	0.2834596	0.4325531	0.5000000
KENNEDY	4.560555e-04	0.5616853	0.7449995	0.9054510	0.2988641	0.4549130	0.4995500
LOS MARTIRES	0.000000e+00	0.5712640	0.7275510	0.8874418	0.3393969	0.4442258	0.4978632
PUENTE ARANDA	3.021027e-04	0.5801791	0.7520670	0.9243955	0.3310811	0.0000000	0.0000000
RAFAEL URIBE URIBE	4.646840e-04	0.5586052	0.7370987	0.9054156	0.3175651	0.4534950	0.5000000
SAN CRISTOBAL	4.127967e-04	0.5656856	0.7446725	0.8994898	0.3640313	0.4463457	0.4973545
SANTA FE	2.041126e-03	0.5680979	0.7300607	0.8773359	0.3282643	0.4447133	0.5000000
SUBA	2.942264e-04	0.5637876	0.7436055	0.9075351	0.3402679	0.4469239	0.5000000
TEUSAQUILLO	0.000000e+00	0.5716270	0.7364094	0.8943676	0.0000000	0.4166667	0.0000000
TUNJUELITO	0.000000e+00	0.5686275	0.7379609	0.9200742	0.0000000	0.4375000	0.0000000
USAQUEN	9.157509e-05	0.5560776	0.7463202	0.8953661	0.3428158	0.4367492	0.5014493
USME	2.693900e-03	0.5629002	0.7245153	0.8882781	0.3093226	0.4449062	0.5015540
NO REGISTRO	0.000000e+00	0.5358172	0.7138401	0.8518175	0.3492325	0.4458019	0.4965753

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

En total son 7 categorías que buscan explicar en cada localidad como es el comportamiento del indicador, teniendo en cuenta que la operación de la creación de la tabla de contingencia (tabla cruzada), es el promedio del indicador para el cada de registro por localidad.



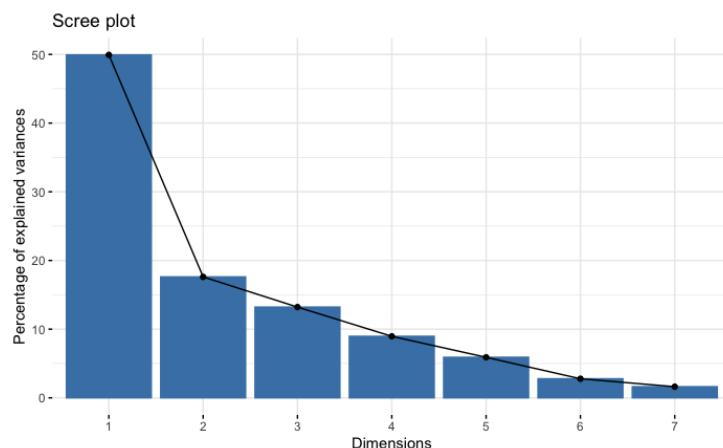
En la localidad de "Ciudad Bolívar" y "Usme" en promedio se presentan manzanas con niveles de Alto Déficit, las localidades "Tunjuelito" y "Teusaquillo" presentan en promedio las menores concentraciones de niveles en Déficit Medio. Los niveles de Alto Superávit se destacan para las localidades Puente Aranda, Engativá y Tunjuelito. Es de notarse, que para la localidad de "Puente

Ilustración 25. Mapa de calor de la tabla de contingencia del indicador DCSP

Aranda", el promedio del valor de indicador se distribuyen con mayor magnitud en las categorías "Superávit", "Medio Superávit" y "Alto Superávit", a su vez es la localidad que menor niveles promedio del indicador tiene en la categoría "Déficit".

Análisis de componentes principales (PCA) exploratorio para el indicador: DCSP.

Gráfico 14. Scree plot: PCA gráfico de barras de componentes principales.



La evaluación de las cantidades de componentes necesarias para explicar un porcentaje específico de la varianza de los datos se presenta gracias al "Scree plot", que indica que con dos componentes es posible explicar más del 68% (50% primer componente + 18% segunda componente) de la variabilidad de los datos y que el otro 32% de información es disipada o perdida. Por consiguiente:

Aplicando el método circular y gráfico PCA, asumiendo una perdida de información de exactamente 32.5% se obtiene los siguientes resultados:

Diagrama 5. PCA método gráfico de tablas cruzadas

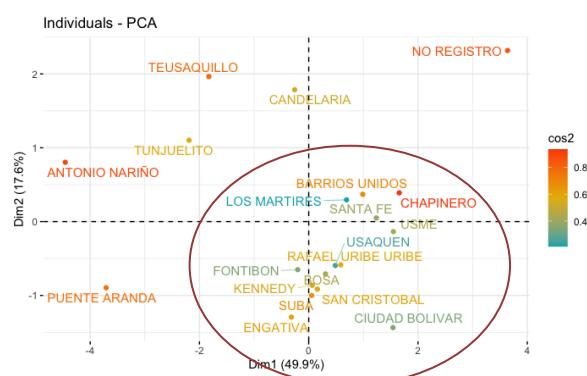
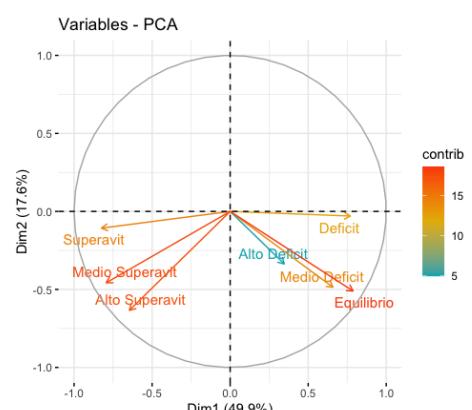


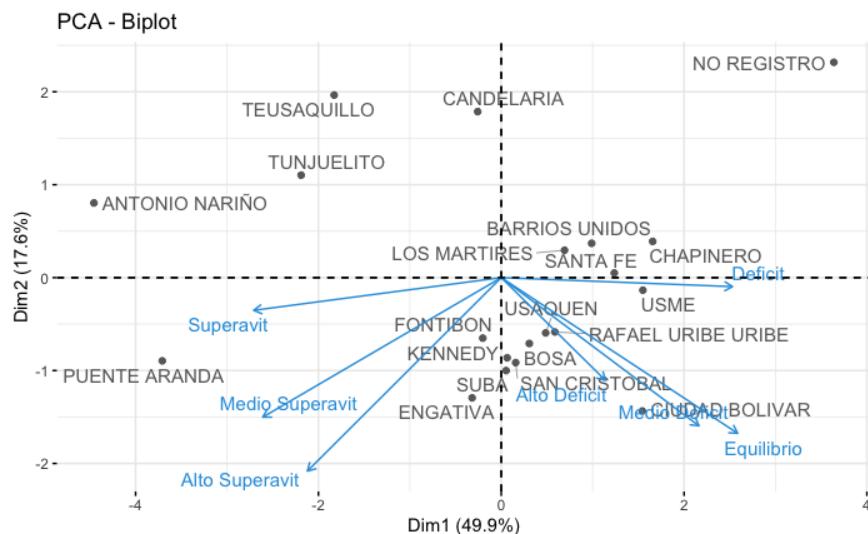
Diagrama 4. PCA método circular de tabla cruzada.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Se muestran en los **Diagramas 5 y 4**, que para dos componentes principales que buscan explicar el comportamiento de la variabilidad en un 67.5%, la mayor cantidad de localidades delimitadas por elipses de contorno rojos se encuentran ubicadas muy próximas a los cuadrantes I, III y IV, esto indicando que las direcciones en términos de categorías que claramente se aprecian tienen permanencia en el cuarto cuadrante IV. Los componentes principales en el método circular, correspondiente a las abscisas en dirección negativa de la componente presentan perfiles similares, es decir se encuentra una asociación de valores por localidad que comparten distribuciones de medición del indicador inter-correlacionados conforme más agrupados se encuentren en el **Diagrama 17**, quiere decir que las mediciones promedio del indicador para todas las categorías son similares o de resultados parecidos, este fenómeno permite explicar que las Localidades como "Teusaquillo", "Tunjuelito", "Candelaria", "Antonio Nariño" y "Puente Aranda" se alejan de los perfiles en términos de resultados, esto quiere decir que el comportamiento del indicador para todas las categorías es particular para estas localidades y dependiendo de las direcciones en las que se encuentren en el plano, pueden referir de algún tipo de tendencia relativa entre localidades.

Diagrama 6. Diagrama Biplot Análisis de componentes principales.

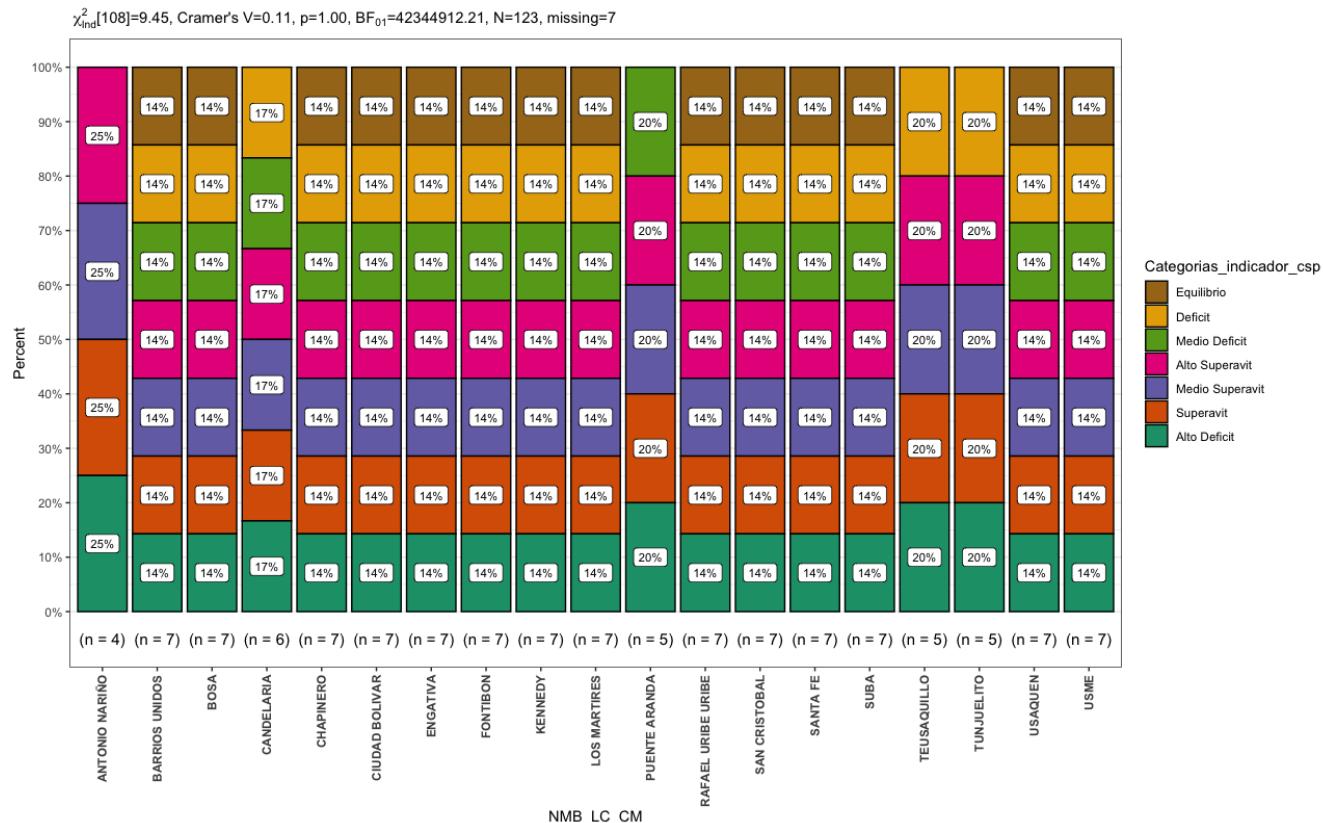


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

Los gráficos cruzados de barras permite visualizar de forma práctica, los porcentajes de participación conforme a las manzanas registradas de la base de datos, respectivamente:

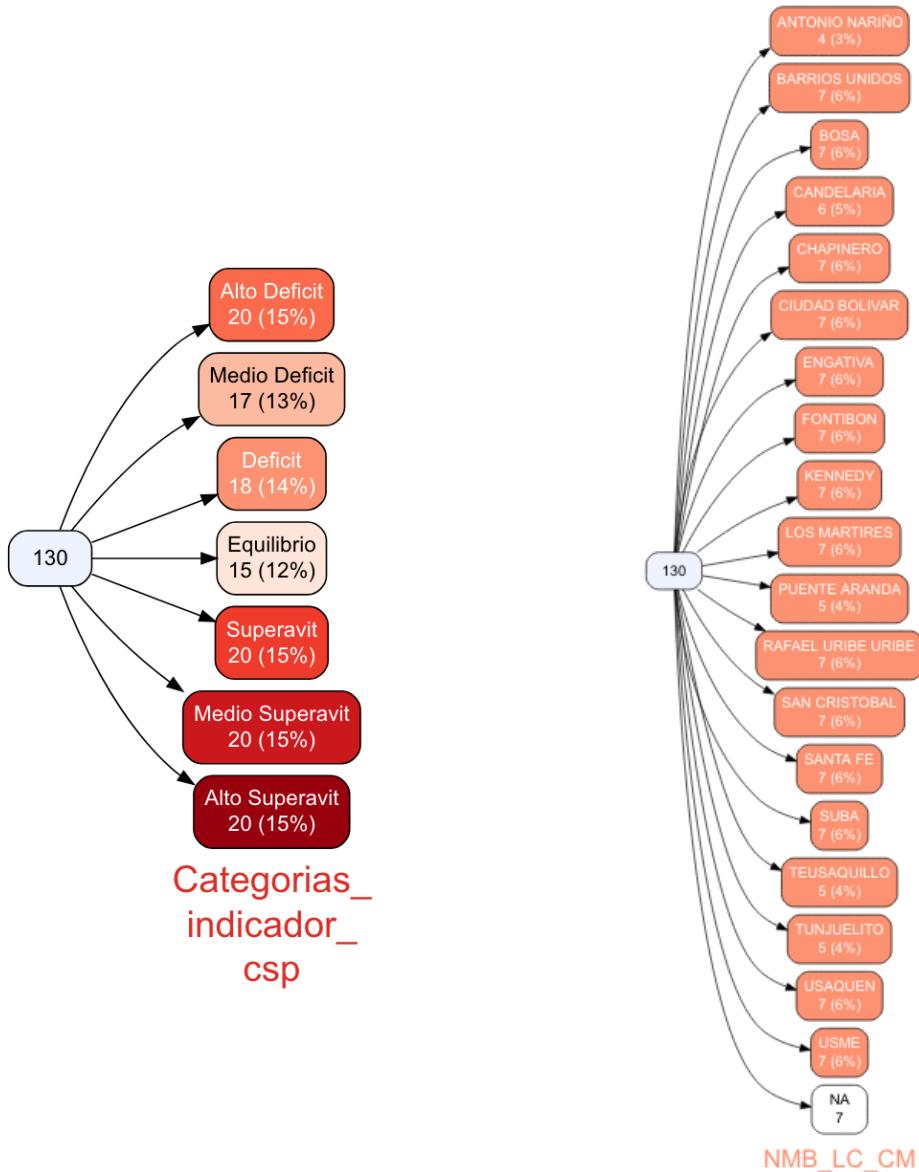
Análisis de tablas de contingencia para el indicador : "DCSP"

Gráfico 15. Gráfico de tabla cruzada del indicador DCSP.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

El anterior esquema presenta las participaciones de las categorías por localidad, por tanto es de notarse que de las 20 localidades, 14 localidades presentan los mismos resultados, entendiéndose que 1 de cada 7 manzanas se encuentra en cualquier tipo de categoría para ese grupo de localidades, adicionalmente se contempla que las localidades "Teusaquillo" y "Tunjuelito" presentan las mismas participaciones de 5 categorías, estas ascendiendo que 1 de cada 5 manzanas en promedio pueden encontrarse entre las categorías "Alto Déficit", "Déficit", "Superalvit", "Medio Superávit" y "Alto Superávit".

Diagrama 7. Diagrama de árbol de participación por categoría y por localidad DCSP.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde RStudio, Observatorio Inmobiliario Nacional DANE.

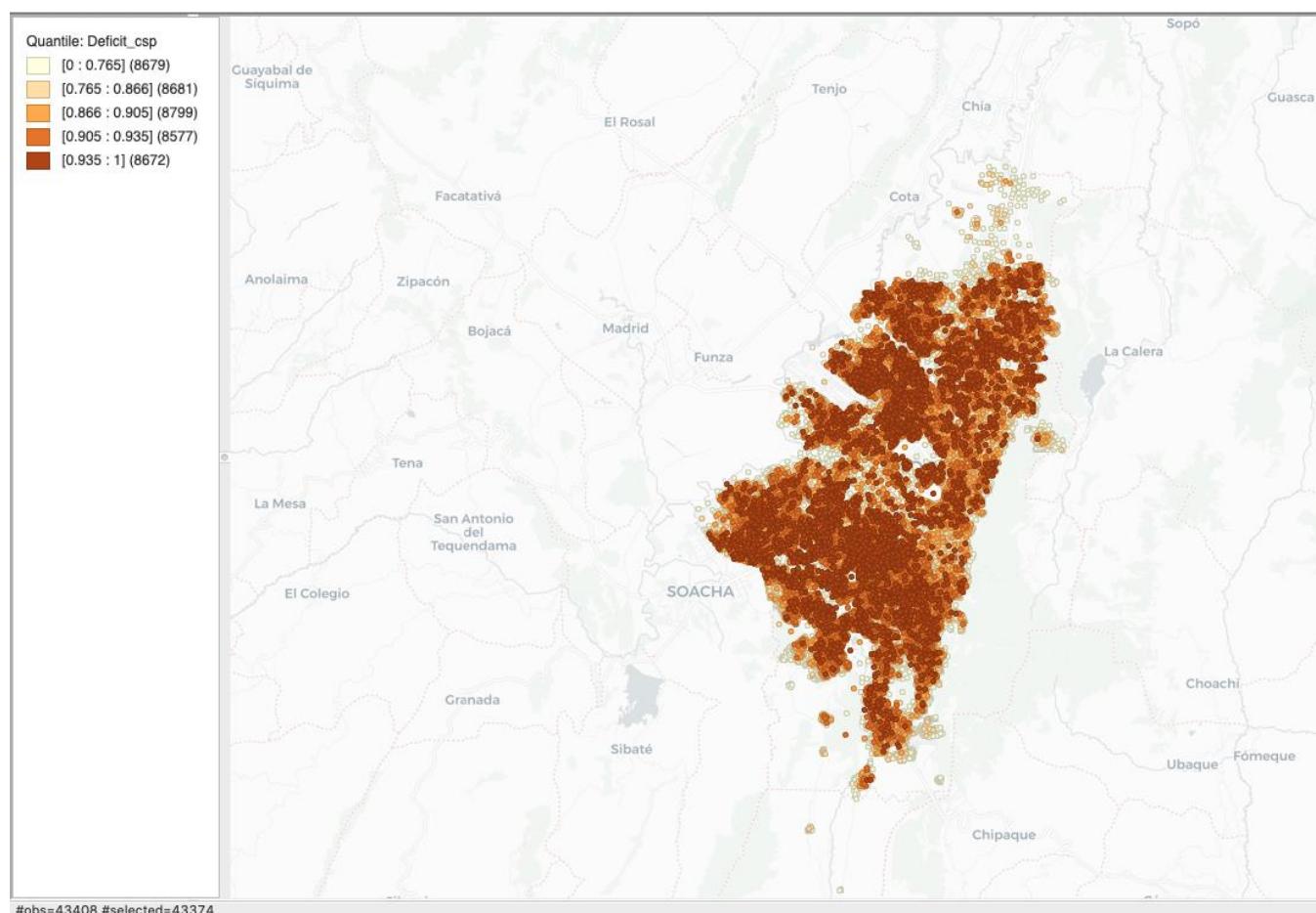
Los porcentajes para cada categoría del en promedio por manzana agrupados por localidad el indicador varia gradualmente, para los 130 valores de los registros del indicador a priori de la construcción de la tabla de contingencia (cruzada), la mayor concentración esta dada a nivel general por registros ubicados en las categorías: "Alto Superávit", "Medio Superávit", "Superávit" y "Alto Déficit" con un 15% para cada una, seguido de "Déficit" con un 14% y "Medio

Déficit" con un 13%, lo que equivale a decir aproximadamente que el 88% de las manzanas se encuentran en estas categorías y el 12% restantes corresponden a la categoría "Equilibrio". (Chavez, 2021)

Construcción de mapas temáticos de la medición del indicador "Déficit de cobertura de servicios públicos": DCSP.

Análisis de localidades con mediciones elevadas del indicador :

Ilustración 26. Mapa de distribución geoespacial del indicador DCSP en Cuartiles. (5)

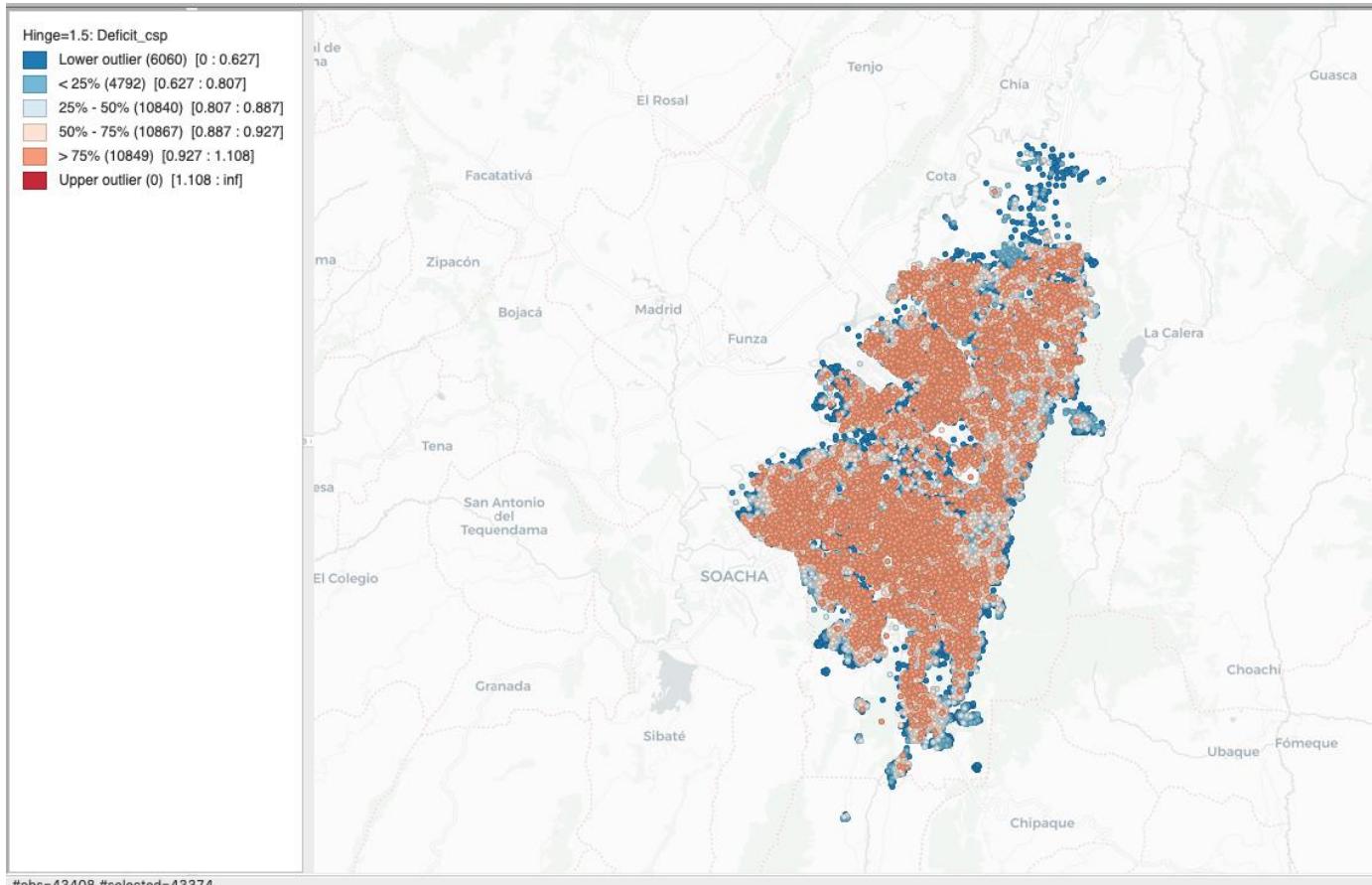


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

Los mapas de cuartiles representan la clasificación del valor del indicador en la ciudad de Bogotá, para un volumen de manzanas iguales entre categorías de orden ascendente el superávit habitacional de entorno se pronuncia de forma uniforme en la ciudad a excepción de la zona central. Una clasificación que contempla los rangos intercuartílico de la distribución de

los datos, se muestra en la **Ilustración 27**, donde se contempla la presencia de Outliers por encima y por debajo del rango intercuartílico.

Ilustración 27. Mapa de distribución geoespacial del indicador DCSP en Cuartiles. (1.5 Rango Intercuartílico). (Gráfico de clasificación intercuartílico)

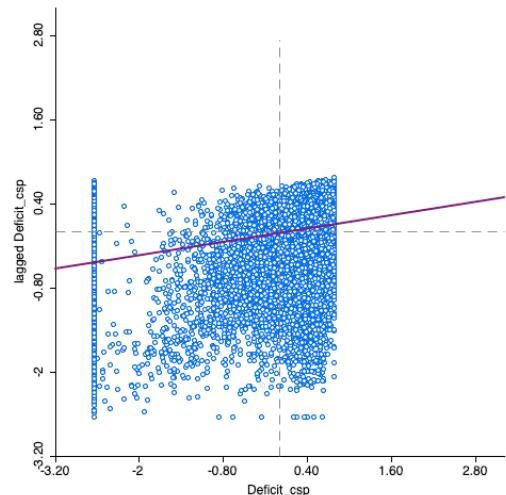


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

Análisis de localidades con mediciones estadísticamente significativas del indicador :

Gráfico de dispersión de Morán:

Gráfico 16. Gráfico de dispersión del índice de Morán : DCSP.

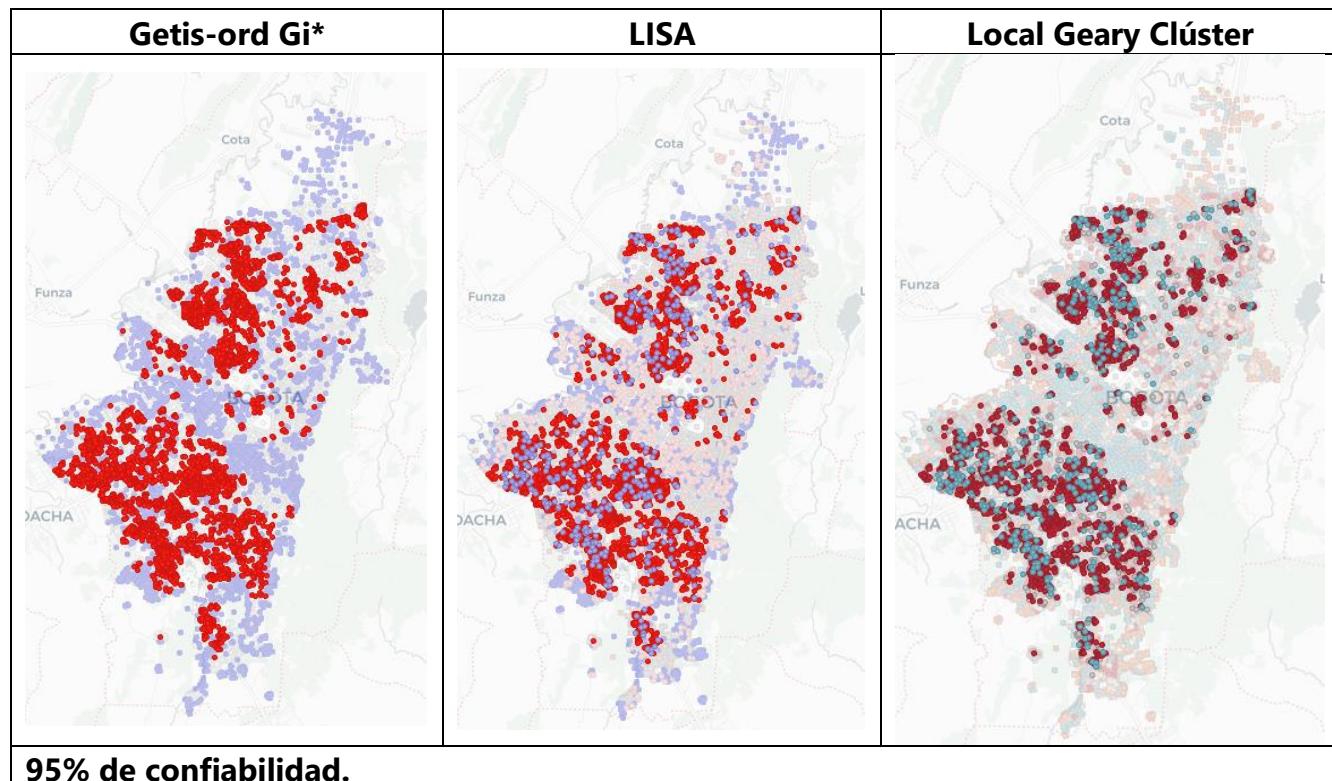


Debido a que el índice de morán se encuentra expresado en una magnitud de 0.157 aproximadamente 15.7%, los datos en su totalidad presentan patrones completamente aleatorios conforme a una distribución espacial. De acuerdo al gráfico de Morán existen significativas cantidades de manzanas que se ubican en los cuadrantes atípicos.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, Observatorio Inmobiliario Nacional DANE.

Comparación de métodos de auto-correlación espacial:

La comparación de métodos de auto-correlación espacial es esencial para seleccionar el método con mayor adecuación que permita explicar los fenómenos, debido que a mayor diferencia de fuentes procedimentales los métodos que expresan resultados similares se determinan asertivos. Una vez implementado el método, se analizan las manzanas que reflejan de alto grado de significancia o que sus valores no dependen de sucesos aleatorios, sino que sin importar de los sucesos o condiciones de la población, las zonas en las que el indicador en una manzana comparte un valor de cobertura completa con las manzanas del sector se mantienen. A continuación se presentan tres mapas de clústeres, cada uno con un método de auto-correlación espacial.

Tabla 18. Mapas de clústeres del Indicador DCSP con grado de verosimilitud del 99%.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

Se analizan las leyendas de los métodos de acuerdo a las cantidad de manzanas integradas para cada categoría, respectivamente:

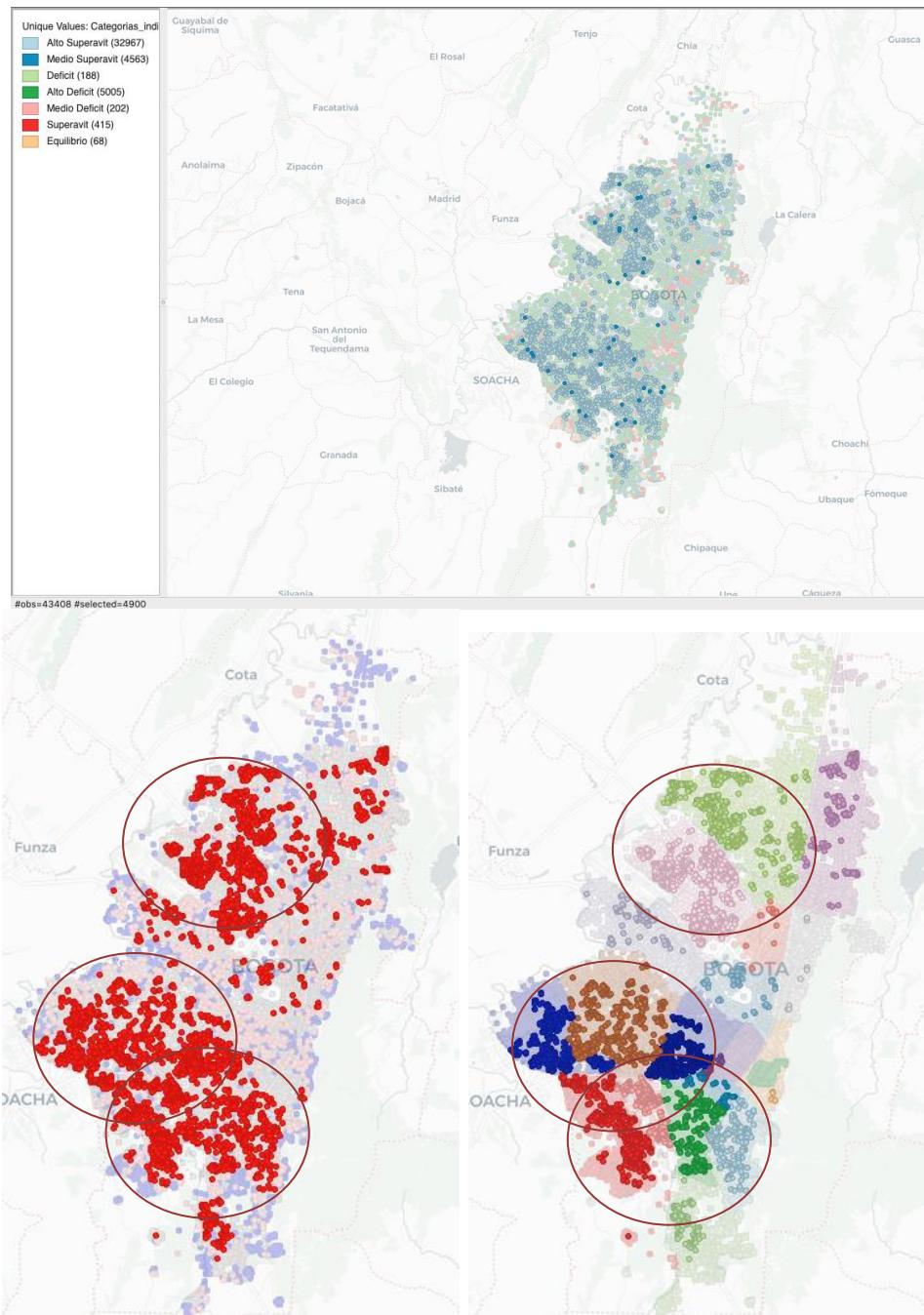
Tabla 19. Leyendas de Mapas de significancia del DCSP.

Getis-ord Gi*	LISA	Local Geary Cluster
<ul style="list-style-type: none"> ■ Not Significant (33952) ■ High (5368) ■ Low (4088) 	<ul style="list-style-type: none"> ■ Not Significant (33952) ■ High-High (4900) ■ Low-Low (2216) ■ Low-High (539) ■ High-Low (1801) 	<ul style="list-style-type: none"> ■ Not Significant (31625) ■ High-High (6622) ■ Low-Low (2258) ■ Other Positive (625) ■ Negative (2278)

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

De acuerdo a la **Tabla 19**, a diferencia de los métodos de "Gi*" y "LISA", el método "LGC" tiene alrededor de 6622 manzanas que son de contigüidad "Alto-Alto" ("High-High"), siendo 1.35 veces mayor que la cantidad expresada por el método LISA, 1.23 veces mayor que el método Gi*, por lo que este método tiene mayor número de manzanas que no están siendo consideradas HotSpots por otros métodos careciendo de precisión.

Ilustración 28. Mapas de clústeres del método LISA vs Mapa único de localidades vs Mapas único de Categorías del Indicador DCSP.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde GeoDA, OIN DANE.

En el 95% de los casos existe mayor población con cobertura en servicios públicos y con todos los servicios en las localidades "Suba", "Teusaquillo", "Engativá", "Puente Aranda", "Kennedy", "Bosa", "Rafael Uribe Uribe" y "Tunjuelito", apreciado en la **Ilustración 28** que presenta los mapas de clústeres. Aquellas zonas en las que se evidencia presencia de manzanas como vectores en puntos rojos precede de la existencia de la posibilidad del evento de encontrar manzanas con "Alto superávit" rodeadas de manzanas en "Alto Superávit".

Ilustración 29. Mapas condicionales bivariado: DDPCSP vs DCSP



En el mapa condicional bivariado de la **Ilustración 29**, se detalla que la relación que guarda el indicador DDPCSP con el indicador DCSP, muestra la representación de valores atípicos, en el que para niveles cercanos a 1 del indicador DDPCSP y para valores oscilantes de 0,915 del indicador de cobertura en servicios públicos (DCSP), coinciden 2886 manzanas ubicadas al norte y centro de la ciudad para ambos valores que están por encima del rango intercuartílico, esto indicando que para la combinación de valores de ambos indicadores expuestos están ubicados en las zonas mencionadas.

Ilustración 30. Análisis de las manzanas con mediciones elevadas del indicador : DCSP

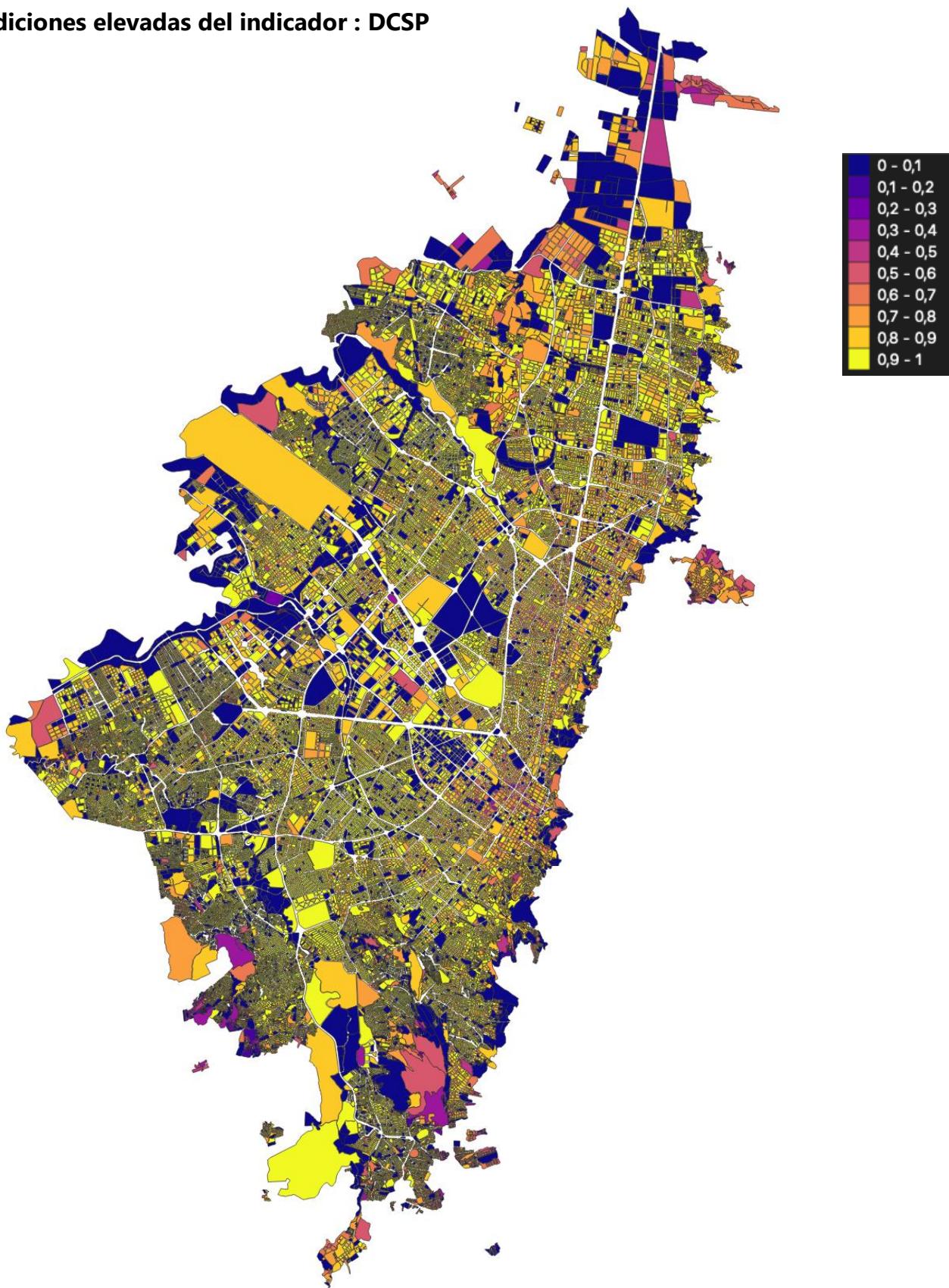
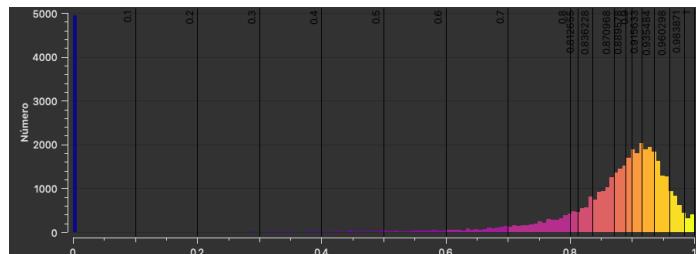


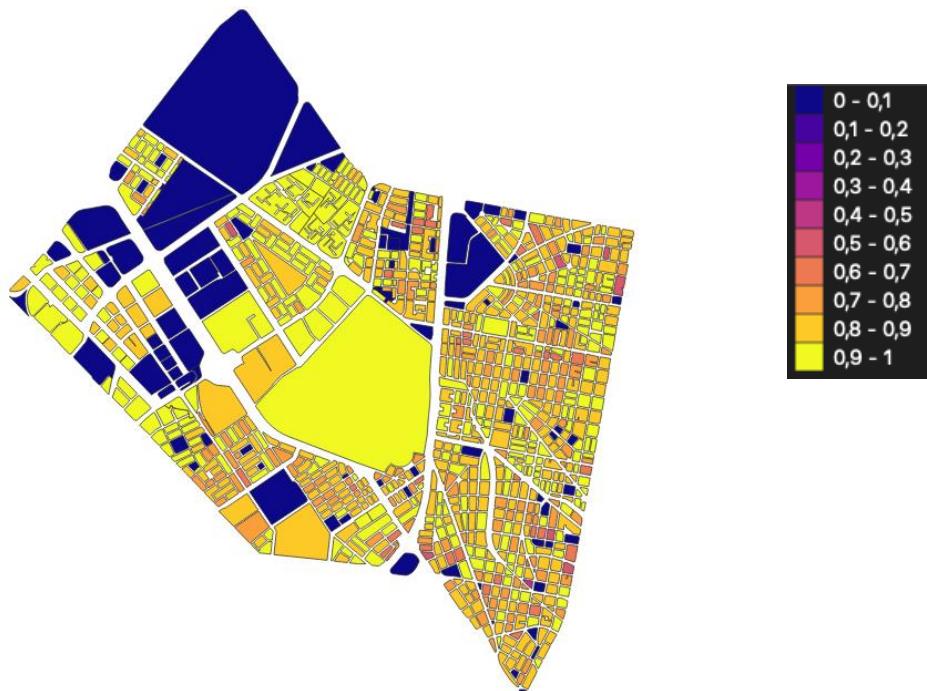
Gráfico 17. Histograma del indicador DCSP en la ciudad de Bogotá DC

En la clasificación del indicador por manzana en la ciudad de Bogotá DC, se presenta la **Ilustración 30**, que muestra la distribución de las manzanas conforme a sus colores dados por la clasificación una alta presencia de manzanas con colores para valores entre 0.8 y 1. En el **Gráfico 17**, se contempla la existencia de muchos registros de manzanas con altas de áreas que se encuentra en "Alto Superávit", la población en cuestión tiene más de 4 servicios.

Análisis de la localidad de Teusaquillo con mediciones del indicador :

En la localidad de Teusaquillo, se presentan manzanas con valores dentro de las escalas del [0,1], con alta presencia de datos en "Alto Déficit"[0,0.2] ubicados al Norte de la localidad, en el extremo "Noroeste" de la ciudad en los polígonos más grandes.

Ilustración 31. Mapa localidad Teusaquillo con distribución geoespacial del indicador DCSP.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE.

En el barrio de “La Esmeralda”, es de notarse que la mayoría de manzanas se ubican en el rango de 0.8 a 1 valores del indicador. Existen manzanas que se encuentran en un tipo de “Alto Superávit” tal como lo son las manzana distribuidas con color naranja con medición equivalente mayor a 0.8 pero también se encuentra la manzana con medición equivalente a 0, clasificada dentro de la categoría “Alto Déficit”, la población en esa manzana presenta no registro de tener ningún tipo de servicio.

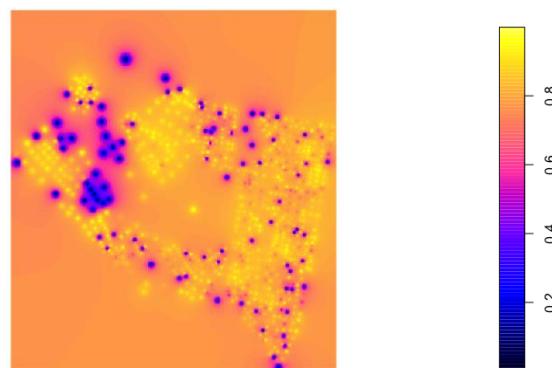
Ilustración 32. Mapas de selección de manzanas en el barrio “La Esmeralda”



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE.

En la **Ilustración 33** se presenta la interpolación con Kriging con la determinación de valores de acuerdo a la covarianza del indicador, contemplando los valores de manzanas de la localidad de Teusaquillo en Bogotá DC. Es de compararse las zonas de esta población considerando un escala de máximos y mínimos conforme a la medición del indicador en esta segmentación geográfica.

Ilustración 33. Mapa de interpolación kriging del indicador DCSP.

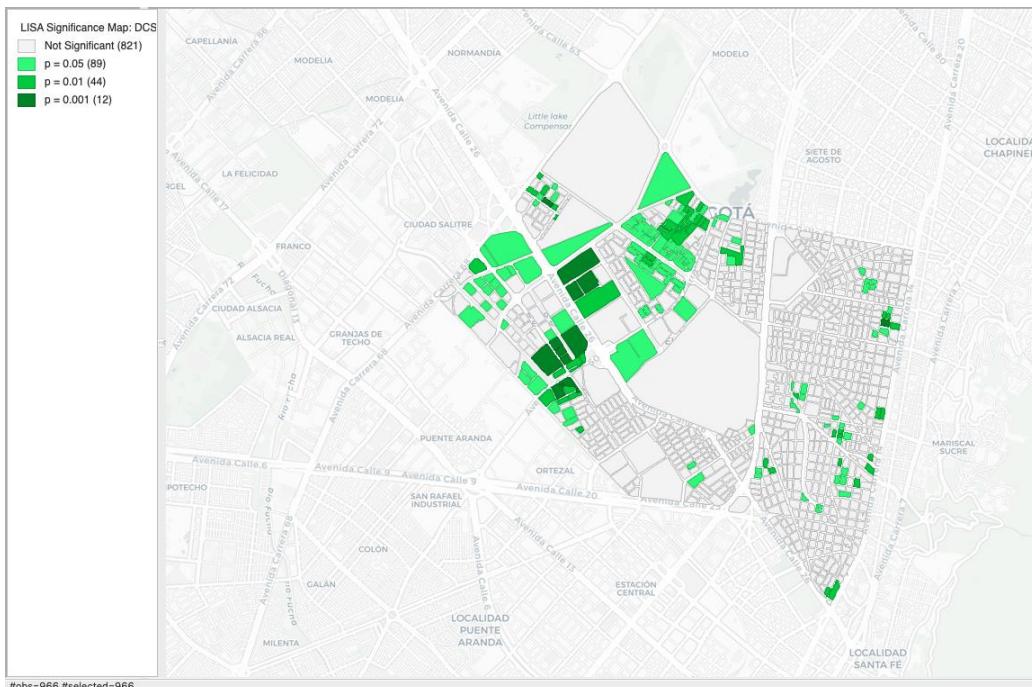


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Se contemplan un gran cantidad de vectores en los que los indicadores presentan altas magnitudes comparados con las manzanas en la misma localidad, estos valores por encima del promedio podrían ser considerados en "Superávit".

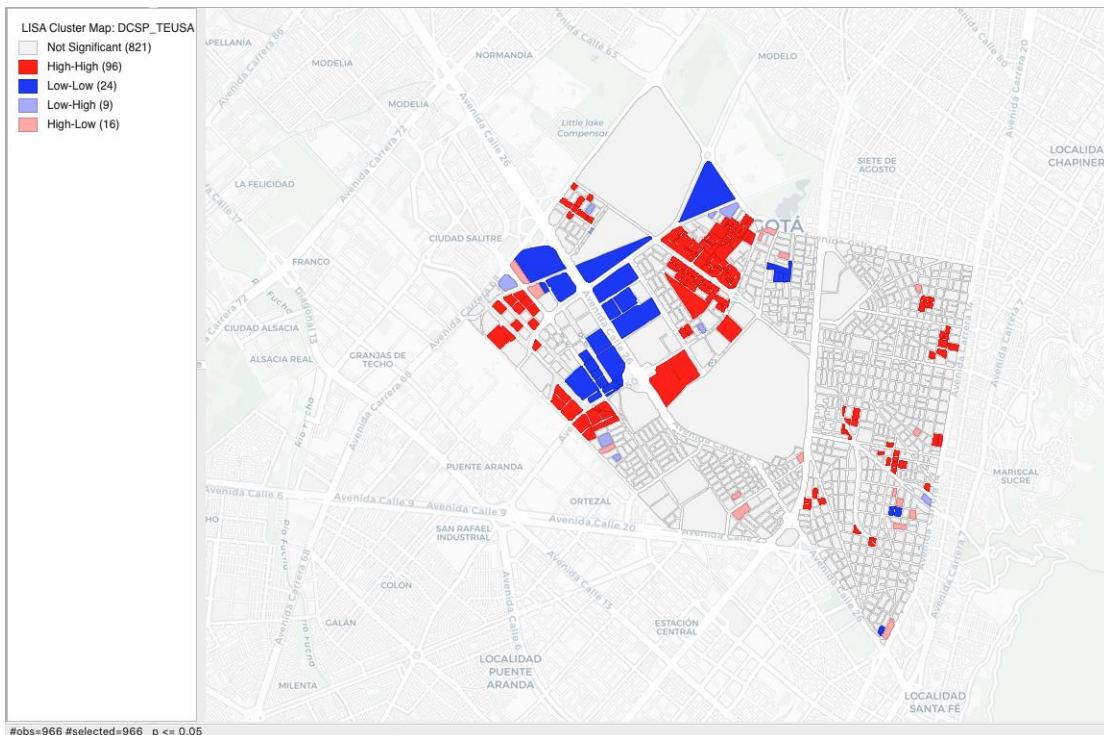
Análisis prospectivo de la localidad de Teusaquillo con mediciones del indicador : Auto-correlación espacial.

Ilustración 34. Mapa de significancia estadística método LISA : Localidad Teusaquillo.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de GeoDA, Observatorio Inmobiliario Nacional DANE. Se presentan 12 manzanas con elevada significancia estadística, estas distribuidas en diferentes barrios señalando la fuerte existencia de la posibilidad de encontrar efectivos del indicador en distintas zonas de la localidad que contengan un déficit del índice de servicios, la población con cobertura a servicios públicos que tenga a completitud la cantidad de servicio.

Ilustración 35. Mapa de clúster método LISA : Localidad Teusaquillo.

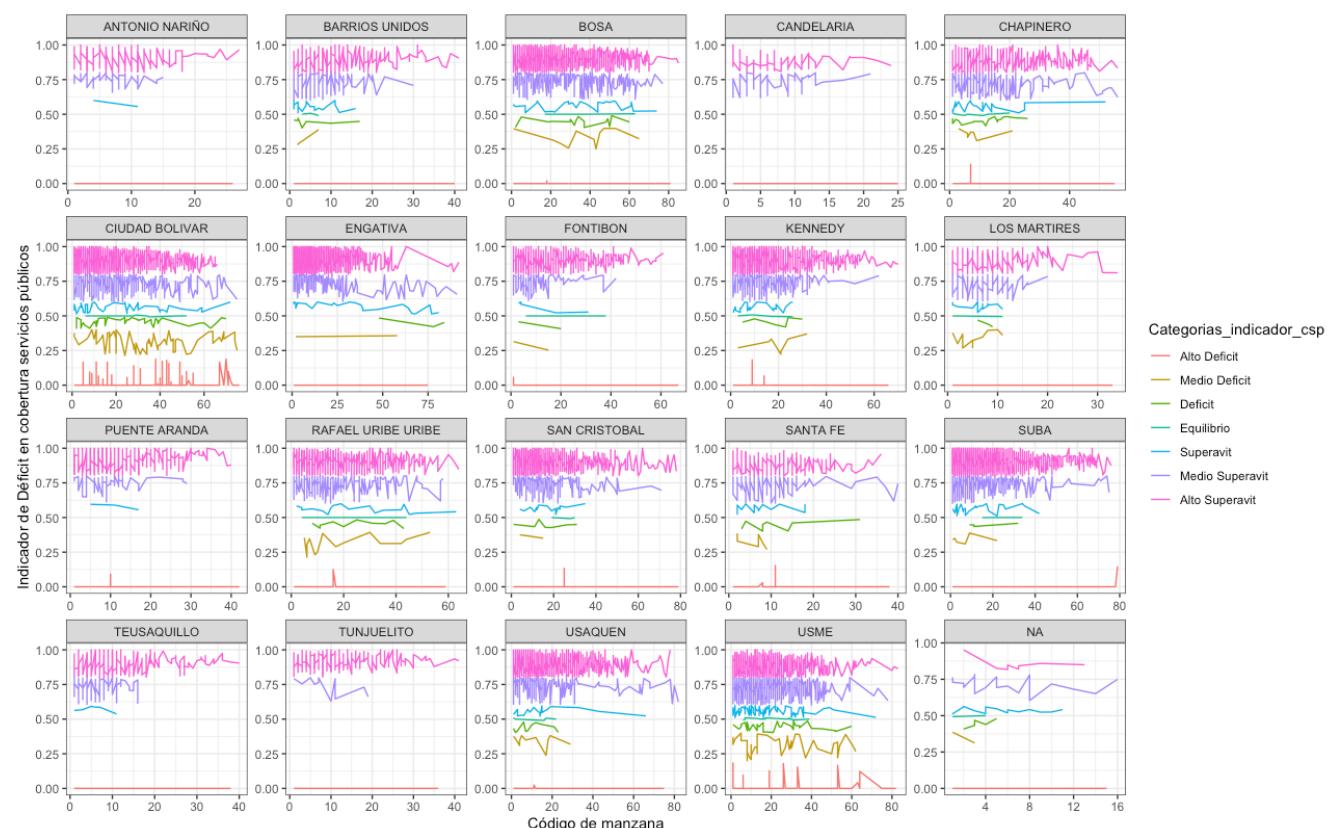


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de GeoDA, Observatorio Inmobiliario Nacional DANE.

El método Local Morán demuestra que el 95% de los casos las manzanas agrupadas en la que la población con cobertura a servicios públicos predomina en las manzanas de color rojo, y en las manzanas de color azul hay mayor agrupación de manzanas con concentración poblacional con valores bajo del indicador comparado en la misma localidad para Teusaquillo hace referencia a las manzanas que se encuentran en equilibrio, es decir que para los barrios "Pablo VI" y el barrio "La Esmeralda", se encuentran clústeres en los que las manzanas todas tienen niveles altos de cobertura a servicios públicos.

Comparando los valores del indicador para cada localidad se encuentra que para Teusaquillo, el comportamiento de los datos para esta variable se encuentra en un aspecto estacional y con una amplitud elevada en función de la manzana. Para un total de 40 manzanas enumeradas y repartidas en todo el data set, las manzanas del código 1 al 20 presentan un patrón estacional ascendente, en la que los registros del indicador están por encima del 0.5 y menores que 1, no obstante se encuentra a su vez que los datos entre la manzana 20 a la 40 presentan un patrón estacional descendiente pero menos uniforme debido a que los tamaños de estas manzanas varían, adicionalmente se encuentra distribuidas bajo una gran dispersión, los promedios son diferentes para todas las localidades, por ejemplo, es de apreciarse que en la localidad de "Ciudad Bolívar" los registros por debajo de 0.25 en el indicador tienen presencia en todas las manzanas.

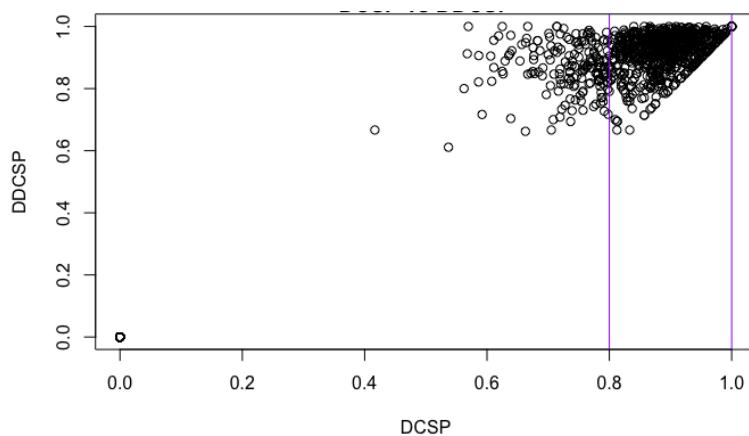
Tabla 20. Gráfico de series para las manzanas en cada localidad : Niveles del DCSP en términos de categorías.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

En la localidad de “Teusaquillo”, la relación entre el indicador DCSP y el DDPCSP, refleja una relación lineal que la mayor cantidad de datos se encuentra entre 1 y 0.8.

Gráfico 18. Gráfico de dispersión DCSP vs DDPCSP

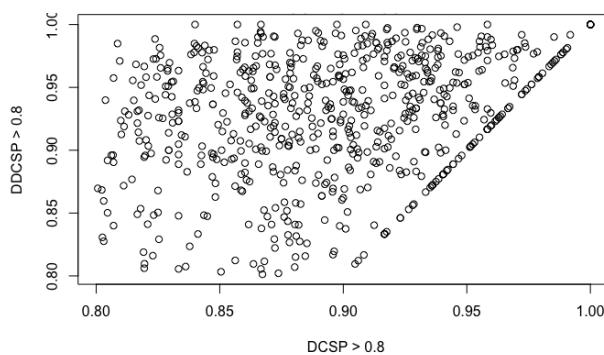


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Analizando la mayor concentración de observaciones se revisan un filtrado de todos los registros de la localidad de “Teusaquillo”, filtrando para estos dos límites mencionados, obteniendo así un acumulado de poblaciones en el diagrama.

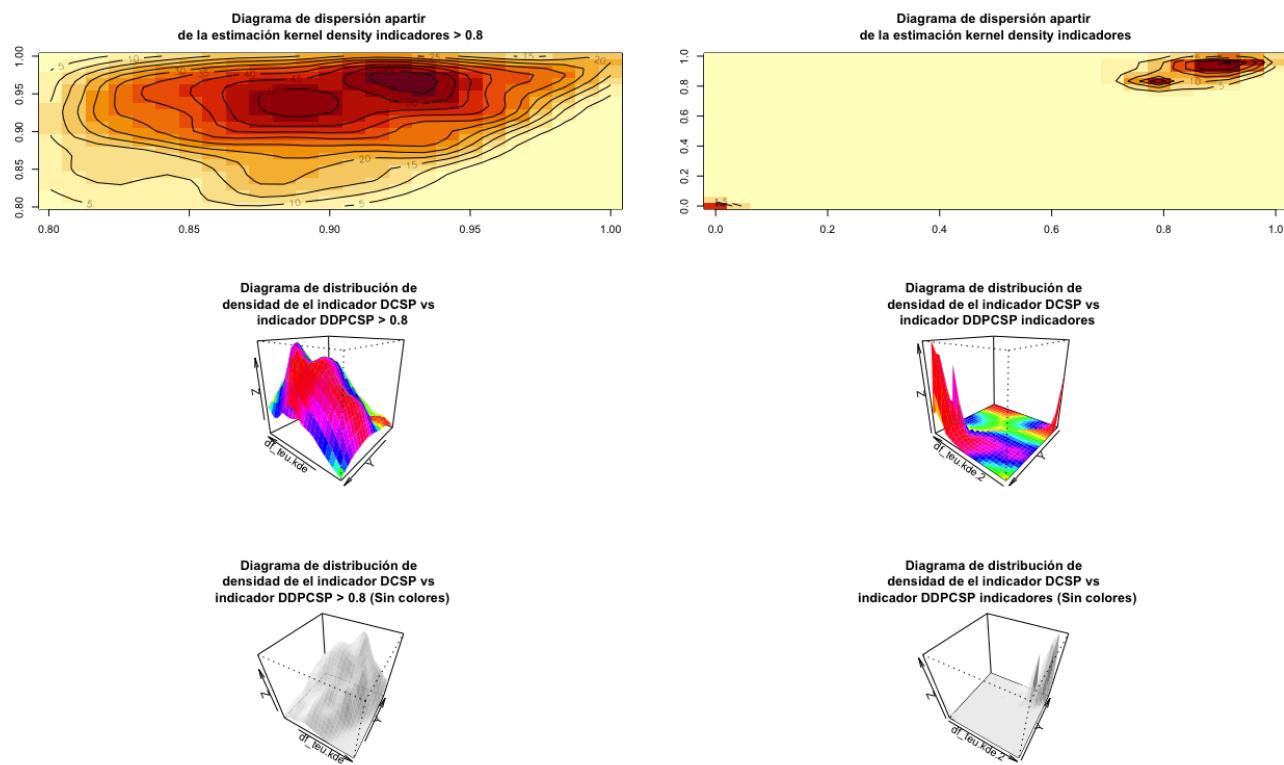
Realizando el filtrado es posible observar que aún se presentan diferentes tendencias y dispersiones, en el caso de los valores de 0.9 a 1 presentan una relación fuertemente correlacionada para algunas manzanas, que indican que si DCSP aumenta DDPCSP también.

Gráfico 19. Gráfico de dispersión filtrado para valores mayores en ambos indicadores a 0.8



Escalando el diagrama de dispersión a una transformación desde el kernel density en tres dimensiones, así siendo permisible estimar la distribución de probabilidad de densidad de estos datos comportados de forma aleatoria y filtrados de acuerdo a la concentración de participación entre los indicadores, respectivamente.

Gráfico 20. Análisis de densidad de kernel: Diagrama dispersión DCSP vs DDPCSP



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

En el filtro del indicador con valores mayores que 0.8 y menores que uno, en su de diagrama de dispersión de densidad **Gráfico 20**, los datos presentan asentamientos de curvas de densidad comparado con el de toda la población de Teusaquillo. El radio con valores más representativos son: El ubicado entre 0.85 y 0.95 de los valores de DCSP; desde 0.90 y 1 para el DDCSP. Es posible inferir a partir del diagrama de distribución de densidad con las fluctuaciones de los datos, que las zonas en las que más se concentran los datos con colores fuertes es en la que las condiciones de la localidad tendrán mucho más peso a la hora de llevar niveles del indicador a una proyección de datos desconocidos o en procesos estocásticos que figuren cambios a través del tiempo pero que del mismo modo devienen un mayor probabilidad de obtener casos parecidos a estas zonas, a grandes rasgos se plantea que es bastante probable

que en una extensión temporal para una porción del conteo de viviendas en la localidad de Teusaquillo, las manzanas tengan más de 5 servicios público y que entre el 90% y el 100% de la población sea de tipo favorecida en un "Alto Superávit".

2.3 Construcción del indicador: "Densidad Habitacional". (DH)

La densidad habitacional hace referencia a la existencia de la satisfacción conforme a las necesidades de alojamiento de los hogares, este valor busca explicar donde se encuentran los excesos de hogares que no son suplidos por las cantidades de viviendas y se entiende como un problema cuando se presentan desequilibrios, debido a que las condiciones habitacionales no son las adecuadas. Las variables a fines a la construcción del indicador DH, son la siguiente:

Tabla 21. Variables a fines a la construcción de los indicadores asociado a la densidad habitacional.

VARIABLE	INDICADOR	CLASE	DESCRIPCIÓN	Categoría original
TVIVIENDA	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas	
TP14_1_TIP	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas tipo Casa	Casa
TP14_2_TIP	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas tipo Apartamento	Apartamento
TP14_3_TIP	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas Tipo cuarto	Tipo cuarto
TP14_4_TIP	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas tipo Vivienda tradicional Indígena	Vivienda tradicional Indígena
TP14_5_TIP	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas tipo Vivienda tradicional étnica (Afrocolombiana, Isleña, Rrom)	Vivienda tradicional étnica (Afrocolombiana, Isleña, Rrom)
TP14_6_TIP	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas tipo Otro (contenedor, carpa, embarcación, vagón, cueva, refugio natural, etc.)	Otro (contenedor, carpa, embarcación, vagón, cueva, refugio natural, etc.)
TP15_1_OCU	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas Ocupada con personas presentes	Ocupada con personas presentes

TP15_2_OCU	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas Ocupada con todas las personas ausentes	Ocupada con todas las personas ausentes
TP15_3_OCU	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas Vivienda temporal (para vacaciones, trabajo etc.)	Vivienda temporal (para vacaciones, trabajo etc.)
TP15_4_OCU	DENSIDAD HABITACIONAL	Conteo de viviendas tipo	Conteo de viviendas Desocupada	Desocupada
TP16_HOG	DENSIDAD HABITACIONAL	Conteo de hogares	Conteo de hogares	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de EXCEL, Observatorio Inmobiliario Nacional DANE.

Relacionamiento de hogares y la cantidad de viviendas: DH

Para este apartado se busca medir la incidencia de los hogares en función de cantidad de viviendas de acuerdo a las variables obtenidas en el marco geo-estadístico nacional, a través de un indicador elemental producto de las configuraciones actuales sobre las variables observadas. Con un análisis del déficit cuantitativo de hogares (Densidad de viviendas), se estima la cantidad de viviendas necesarias para la existencia de un equilibrio (relación uno a uno) entre las cantidades de viviendas y el conteo de hogares, en el caso de que el monto de hogares superen a las viviendas se designa entonces como el tipo de "Déficit Cuantitativo".

El nivel agregado por manzana permite estimar la distribución de la densidad habitacional, conforme a la aplicación de la operación de división para las variables "Conteo de Hogares" llamada "TP16_HOG" en el data-set y "Conteo de viviendas" llamada "TVIVIENDA", así mismo se emplea la comparativa como segunda aproximación del indicador una operación de resta, y posteriormente la medición de un análisis factorial exploratorio con el fin de determinar la pertinencia de la aproximación en una variable latente llamada Densidad Habitacional.

Aproximación 1:

Ecuación 17. Primera Aproximación Indicador Densidad Habitacional (Operación Fracción)

$$\text{deficit}_{\text{absoluto}} = \frac{\text{TP16}_{\text{HOG}}}{\text{TVIVIENDA}} = \frac{\text{Conteo de hogares}}{\text{Conteo de viviendas}}$$

Aproximación 2:

Ecuación 18. Segunda Aproximación Indicador Densidad Habitacional (Operación Resta)

$$\text{deficit}_{\text{absoluto}_r} = \text{TP16}_{HOG} - \text{TVIVIENDA} = \text{Conteo de hogares} - \text{Conteo de viviendas}$$

A continuación se muestra un resumen de las aproximaciones de este indicador:

Tabla 22. Resumen estadístico de las variables afines al indicador DH

— Data Summary —————

	Values
Name	data
Number of rows	43408
Number of columns	12

Column type frequency:

numeric	12
---------	----

Group variables None

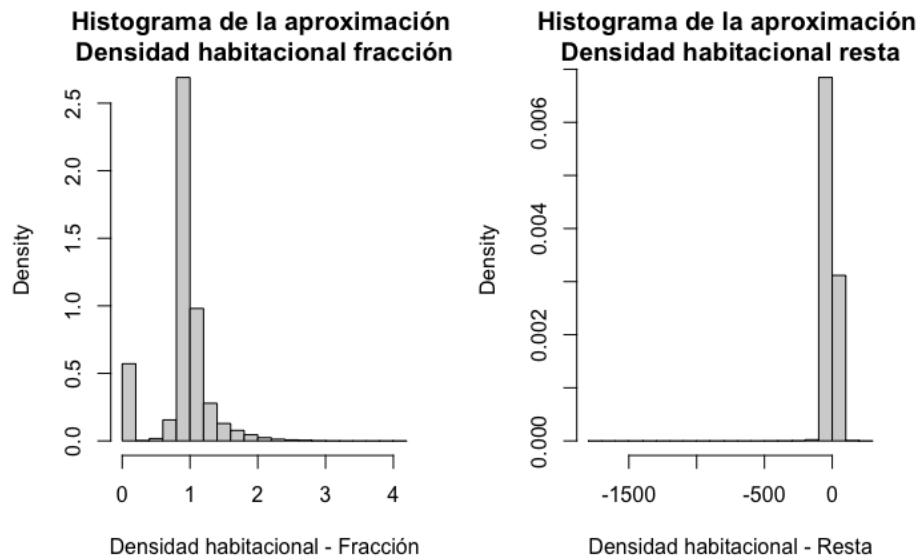
— Variable type: numeric —————

	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1	TVIVIENDA	0	1	58.0	104.	0	17	37	63	4618	
2	TP14_1_TIP	0	1	19.0	36.7	0	5	12	23	1773	
3	TP14_2_TIP	0	1	35.4	94.4	0	2	13	34	4571	
4	TP14_3_TIP	0	1	3.50	7.62	0	0	1	4	231	
5	TP14_4_TIP	0	1	0.00405	0.0730	0	0	0	0	0	
6	TP14_5_TIP	0	1	0.00389	0.0753	0	0	0	0	0	
7	TP14_6_TIP	0	1	0.0384	0.638	0	0	0	0	0	
8	TP15_1_OCU	0	1	53.9	93.0	0	16	35	60	2846	
9	TP15_2_OCU	0	1	0.754	3.98	0	0	0	0	0	
10	TP15_3_OCU	0	1	0.591	2.88	0	0	0	0	0	
11	TP15_4_OCU	0	1	2.72	12.9	0	0	0	0	2	
12	TP16_HOG	0	1	57.8	95.5	0	18	39	66	2863	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Los histogramas de las aproximaciones, se presentan a continuación:

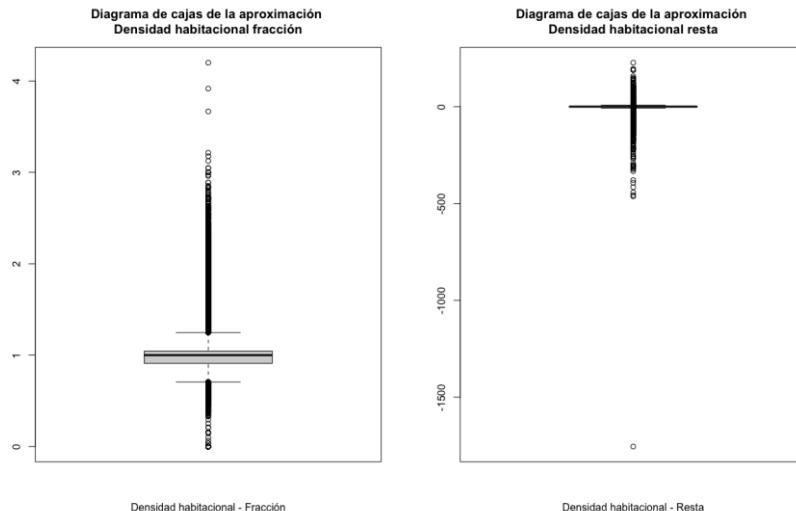
Gráfico 21. Histogramas de las aproximaciones del indicador DH.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

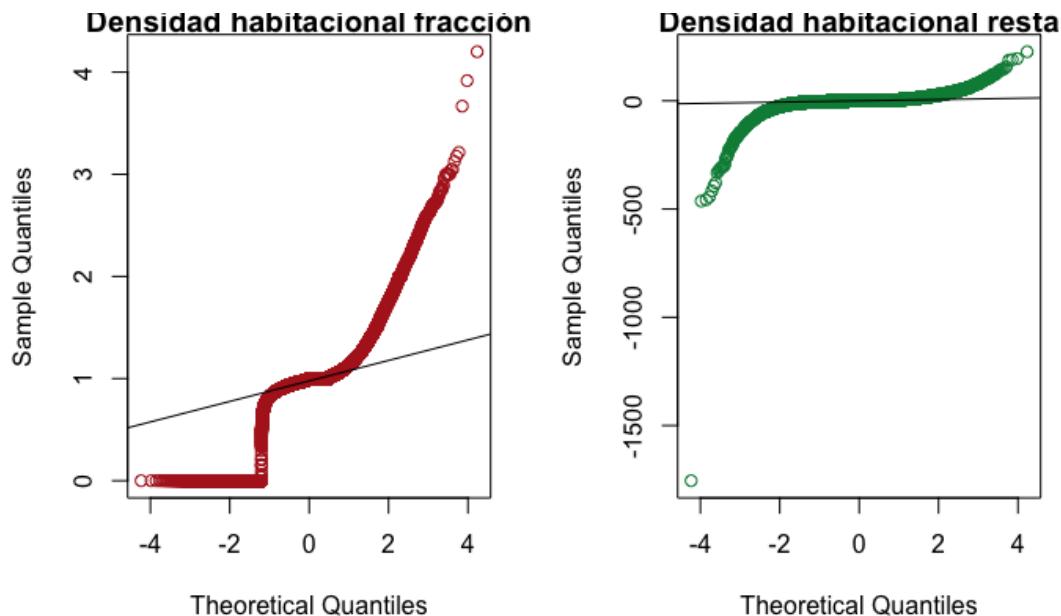
Se observa en el **Gráfico 21** los histogramas, resultando así que para la aproximación con operación “resta” presenta sesgo hacia la izquierda, demostrando una concentración de los datos en el valor 0, lo que nos permite indicar que existen valores atípicos muy por debajo de 1.5 veces el rango intercuartílico; para el caso de la aproximación de fracción se aprecia que la mayor cantidad de datos se concentran en el valor de 1, es posible constatar lo anterior en el siguiente gráfico de “Diagrama de cajas”:

Gráfico 22. Gráficos de cajas de las aproximaciones del indicador DH



Al aplicar los test de normalidad presentados en el **Gráfico 23**, este busca verificar la pertinencia de los datos conforme si su proveniencia es regida bajo una distribución normal o no, debido a que es necesario para los análisis correspondientes, respectivamente:

Gráfico 23. Gráfico de Q-Q de las dos aproximaciones.



```
Warning in ks.test(x = df$Densidad_habitacional_f, "pnorm",
mean(df$Densidad_habitacional_f), :
ties should not be present for the Kolmogorov-Smirnov test
```

```
One-sample Kolmogorov-Smirnov test
```

```
data: df$Densidad_habitacional_f
D = 0.24807, p-value < 2.2e-16
alternative hypothesis: two-sided
```

```
Warning in ks.test(x = df$Densidad_habitacional_r, "pnorm",
mean(df$Densidad_habitacional_r), :
ties should not be present for the Kolmogorov-Smirnov test
```

```
One-sample Kolmogorov-Smirnov test
```

```
data: df$Densidad_habitacional_r
D = 0.2768, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Para ambos supuestos de normalidad, se identifican que ninguna de los dos aproximaciones cumplen los supuestos de normalidad, evidenciados por el test de Kolmogorov cuyo p-valor está por debajo del límite de significancia, por lo que la disposición de los datos y su procesamiento estadístico serían orientando en la estadística no paramétrica.

El análisis de homocesdasticidad no paramétrico (véase la **Ilustración 36**), permite saber si las varianzas son homogéneas para los dos muestras independientes y con un valor de p-value inferior al 5% del grado de significancia se estima que las varianzas son diferentes, así rechazando la hipótesis que plantea que existe igualdad de varianza para las dos variables procesadas.

Ilustración 36. Resultado de la prueba Fligner-Killeen del Homogeneidad de Varianzas.

```
Fligner-Killeen test of homogeneity of variances

data: list(df$Densidad_habitacional_f, df$Densidad_habitacional_r)
Fligner-Killeen:med chi-squared = 17091, df = 1, p-value < 2.2e-16
```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Para conocer de las dos aproximaciones, cual tiene mejores condiciones estadísticas para explicar el indicador, se aplica un "Test de Levene" en función de las localidades de la ciudad, así permitiendo conocer el ajuste de acuerdo al valor de Fisher-Value, respectivamente:

Ilustración 37. Resultado de la prueba Levene del Homogeneidad de Varianzas para las localidades.

```
Warning in leveneTest.default(y = y, group = group, ...) :
  group coerced to factor.
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value    Pr(>F)
group     18  47.747 < 2.2e-16 ***
                 43312
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Warning in leveneTest.default(y = y, group = group, ...) :
  group coerced to factor.
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value    Pr(>F)
group     18  57.872 < 2.2e-16 ***
                 43312
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El test representado en la **Ilustración 37**, encuentra diferencias significativas entre las varianzas de las localidades de la ciudad y a partir del valor de "Fisher", se considera la opción a priori de seleccionar como indicador explicativo la aproximación operación "fracción", debido a que su "F-Value" es inferior a la operación del "diferencial" comparado con la aproximación de la operación "resta".

Se procede a explicar bajo un análisis factorial exploratorio, la relación de las variables observadas con unas posibles variables latentes (factores) que permitan ponderar cada aproximación de acuerdo a sus condiciones factoriales, teniendo en cuenta las variables discriminadas para este indicador para posteriormente emplear el Análisis Categórico y Geoestadístico del indicador a considerar.

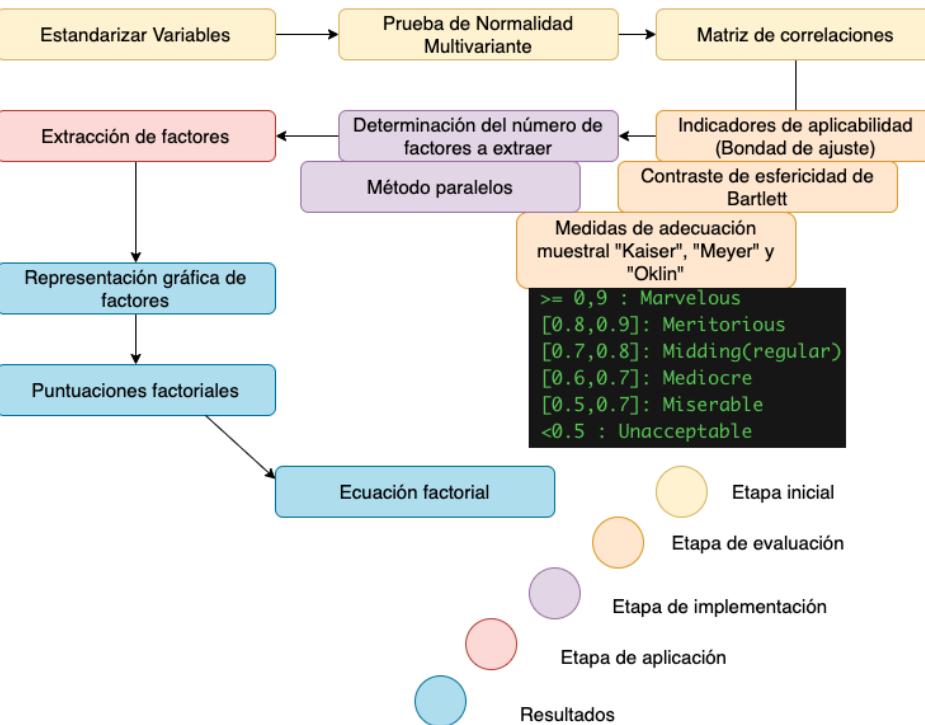
Análisis factorial exploratorio: (AFE)

El análisis factorial exploratorio es una técnica multivariante de inter-dependencia de los datos aplicadas a las correlaciones de las variables métricas, los factores buscan explicar la relación entre variables latentes con respecto a las variables observadas.

Debido a que el comportamiento esta sujeto a cambios drásticos en la composición de los datos referido de forma más clara a la fuente de extracción de los datos, se incorpora un término de error o residuo que buscar integrar parte del comportamiento de las variables latentes por factores en función de la variancia específica, varianza residual o especificidad. En este análisis factorial exploratorio, se busca realizar agrupaciones de variables en indicadores (variables latentes) que permitan explicar de acuerdo a los datos de cada variable observada la correlación reciproca en factores que reflejan de acuerdo a las distancias de estadísticos con patrones similares.

La secuencia de aplicación de este análisis factorial exploratorio se presenta a continuación:

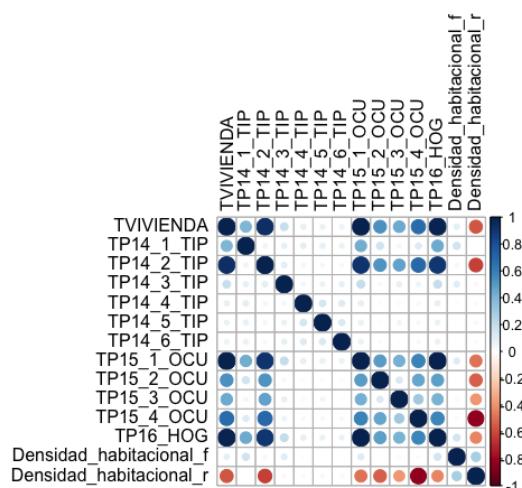
Esquema 1. Metodología del Análisis Factorial Exploratorio



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de draw.io, Observatorio Inmobiliario Nacional DANE.

Partiendo del desarrollo desde la creación de matriz de correlaciones, se presentan los siguientes resultados:

Gráfico 24. Matriz de correlaciones de las variables afines al indicador DH.



Existen clústeres dentro de la matriz de correlación, para las mayores correlaciones presentadas corresponden a las variables de tipo de ocupación que varían para las viviendas ocupadas con personas presentas, las cantidades de hogares, el conteo de viviendas, conteo de viviendas de tipo apartamento y la variable del déficit absoluto bajo la aproximación operacional de resta.

Aplicando el test de correlación presentado en la **Tabla 23**, es posible

conocer la incidencia de esta medida estadística con forme a los datos observados, respectivamente:

Tabla 23. Pruebas de Hipótesis de la medición de correlación global.

$H_0: \text{variable} \rightarrow \text{Correlación} = 0$; (Indica que no existe correlación)

$H_1: \text{variable} \rightarrow \text{Correlación} \neq 0$; (Indica que si existe algún tipo de correlación)

Si H_0 es aceptada, no existe suficiente argumento para explicar AFE.

Si H_0 es rechazada, si existe suficiente argumento para aplicar AFE.

Probability values (Entries above the diagonal are adjusted for multiple tests.)

	TVIVIENDA	TP14_1_TIP	TP14_2_TIP	TP14_3_TIP	TP14_4_TIP	TP14_5_TIP	TP14_6_TIP	TP15_1_OCU	TP15_2_OCU
TVIVIENDA	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP14_1_TIP	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP14_2_TIP	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP14_3_TIP	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP14_4_TIP	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP14_5_TIP	0	0.00	0	0	0.00	0.00	0.00	0	0.02
TP14_6_TIP	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP15_1_OCU	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP15_2_OCU	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP15_3_OCU	0	0.00	0	0	0.00	0.00	0.03	0	0.00
TP15_4_OCU	0	0.00	0	0	0.00	0.00	0.00	0	0.00
TP16_HOG	0	0.00	0	0	0.00	0.00	0.00	0	0.00
Densidad_habitacional_f	0	0.00	0	0	0.01	0.14	0.07	0	0.08
Densidad_habitacional_r	0	0.94	0	0	0.00	0.00	0.00	0	0.00
	TP15_3_OCU	TP15_4_OCU	TP16_HOG	Densidad_habitacional_f	Densidad_habitacional_r				
TVIVIENDA	0.00	0	0	0.00	0.00				
TP14_1_TIP	0.00	0	0	0.00	0.94				
TP14_2_TIP	0.00	0	0	0.00	0.00				
TP14_3_TIP	0.00	0	0	0.00	0.00				
TP14_4_TIP	0.00	0	0	0.05	0.00				
TP14_5_TIP	0.00	0	0	0.28	0.00				
TP14_6_TIP	0.16	0	0	0.27	0.00				
TP15_1_OCU	0.00	0	0	0.00	0.00				
TP15_2_OCU	0.00	0	0	0.27	0.00				
TP15_3_OCU	0.00	0	0	0.00	0.00				
TP15_4_OCU	0.00	0	0	0.00	0.00				
TP16_HOG	0.00	0	0	0.00	0.00				
Densidad_habitacional_f	0.00	0	0	0.00	0.00				
Densidad_habitacional_r	0.00	0	0	0.00	0.00				

$\alpha = 0.05 \rightarrow \text{Alpha (Nivel de significancia)}$

$p - \text{value} > \alpha : \text{Se acepta } H_0$

$p - \text{value} < \alpha : \text{Se rechaza } H_0$

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Se rechaza H_0 , y se concluye que si es posible aplicar un análisis factorial, debido a que la mayoría de variables tienen un valor de p-value igual o muy próximo a cero, a excepción de las variable conteo de viviendas tipo otro (contenedor, carpa, embarcación, vagón, cueva, refugio natural, etc) conocida como "TP14_6_TIP", el conteo de viviendas tipo vivienda tradicional étnica (Afrocolombiana, Isleña, Rrom) denotada como "TP14_5_TIP" y conteo de viviendas con todas

las personas ausentes "TP15_2_OCU", no evidencian estadísticamente algún tipo de correlación con la aproximación operacional de fracción del indicador (Densidad Habitacional Fracción), sin embargo la aproximación operacional de resta (Densidad Habitacional Resta) se correlaciona estadísticamente con todas las variables a excepción de la variable "TP14_1_TIP" que corresponde al conteo de vivienda tipo casa.

En la fase de medición de indicadores de aplicabilidad, se utilizan distintos métodos para corroborar las condiciones actuales de los datos, y determinar con un alto grado de confiabilidad si es posible aplicar el análisis a través del siguiente contraste de esfericidad.

Contraste de esfericidad de Bartlett

Esta técnica se emplea para determinar si a nivel de contigüidad de las variables, las correlaciones teóricas en cada par de variables es nula o no, en caso de ser nula la correlación, existe un argumento suficientemente válido para aplicar el análisis factorial exploratorio.

Por tanto, en este contraste de esfericidad se emplea la siguiente prueba de hipótesis.

Ecuación 19. Aplicación de contraste de esfericidad de Bartlett.

H_0 : Las correlaciones teóricas entre cada par de variables es nula.

H_1 : Las correlaciones teóricas entre cada par de variables no es nulo.

$p - value > \alpha$: no se aplica AFE (Se acepta H_0)

$p - value < \alpha$: si se aplica AFE (Se rechaza H_0)

Bartlett's Test of Sphericity

```
Call: bart_spher(x = data)
```

```
X2 = 4868853.911
```

```
df = 91
```

```
p-value < 2.22e-16
```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Considerando que el $p - value$ es menor que $\alpha = 0.05$, se rechaza la hipótesis nula, afirmando que las correlaciones teóricas entre cada par de variables es nulo por lo que si es aplicable el análisis factorial exploratorio (AFE).

Medida de adecuaciones muestral de Káiser, Meyer y Oklin (KMO)

Esta medida KMO, realiza un estudio de variable por variable que permite determinar si es posible aceptar la participación dentro del proceso a las variables observadas en el AFE, determinando que variables se deberían mantener y que variables deberían ser eliminadas.

Rule of thumbs: Si el KMO de una variables es mayor a 0,7 entonces se acepta que participe, de lo contrario se rechaza, respectivamente con la escala:

Ilustración 38. Criterios de clasificación de la prueba de adecuación muestral KMO.

>= 0,9 : Marvelous
 [0.8,0.9]: Meritorious
 [0.7,0.8]: Middling(regular)
 [0.6,0.7]: Mediocre
 [0.5,0.7]: Miserable
 <0.5 : Unacceptable

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = r)

Overall MSA = 0.5

MSA for each item =

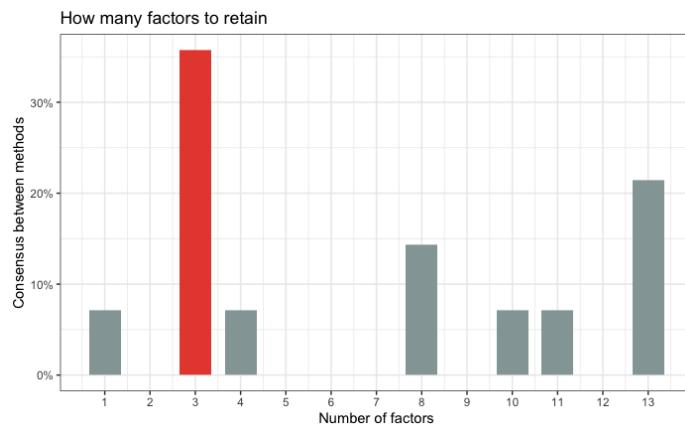
TVIVIENDA	TP14_1_TIP	TP14_2_TIP	TP14_3_TIP	TP14_4_TIP
0.5	0.5	0.5	0.5	0.5
TP14_5_TIP	TP14_6_TIP	TP15_1_OCU	TP15_2_OCU	TP15_3_OCU
0.5	0.5	0.5	0.5	0.5
TP15_4_OCU	TP16_HOG	Densidad_habitacional_f	Densidad_habitacional_r	
0.5	0.5	0.5	0.5	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Aplicando el test de KMO(véase **Ilustración 38**), se obtuvo un puntaje global de 0.5, ubicando a todos los datos en una métrica “Miserable”, este test demuestra que las características estadísticas de las variables expresan una condición bastante precaria en cuanto a un nivel de pertinencia del uso de modelos explicativos conforme a la uniformidad en la medición, no obstante se procede a aplicar el AFE debido a que si bien su condición no es la adecuada, se tiene un puntaje diferente al “Inaceptable”.

En la etapa de implementación, se busca determinar el número de factores a extraer, a partir de diferentes métodos, no obstante también resulta importante determinar el tipo de rotación que se implementaría, el método de estimación y la matriz de estimación.

Para ello se implementa la evaluación del número de factores con una rotación de tipo “varimax”, empleando la función “n_factors” para las variables en estudio y aplicando el algoritmo de máxima verosimilitud, obteniéndose respectivamente.

Gráfico 25. Gráfico de barras de consenso de métodos en función de Factores a extraer.

En el **Gráfico 25**, se presenta el método que indica el consenso entre los métodos, aproximándose a un proporción exacta en que la cantidad de factores que debería ser extraídos y según el gráfico de barras la barra roja indica que la selección debe ser 3 (tres), respectivamente:

Ilustración 39. Reporte del consenso entre métodos AFE.

```
# Method Agreement Procedure:
```

```
The choice of 3 dimensions is supported by 5 (35.71%) methods out of 14 (CNG,  
Optimal coordinates, Parallel analysis, Kaiser criterion, Scree (R2)).
```

Los resultados (**Ilustración 39**) plantean que el 35.71% de los métodos (5 métodos) apoyan la idea que sean seleccionados 3 factores dimensionales, estos 5 métodos de análisis factorial corresponden a: El CNG, Optimal coordinates, Parallel Analysis, el Kaiser criterion y el Scree(R2), cuya procedencia es sobre los métodos del “scree” y “cng”, como se presentan a continuación:

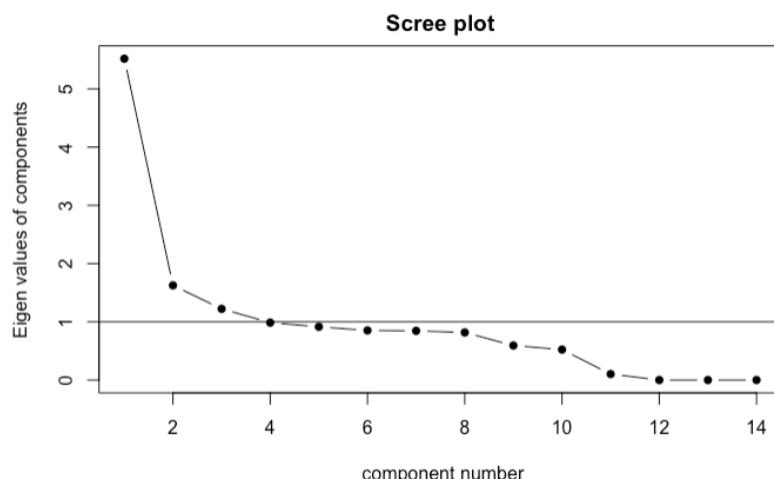
Ilustración 40. Métodos disponible (Total :14) para el Análisis Factorial Exploratorio.

n_Factors	Method	Family
<dbl>	<chr>	<chr>
1	Acceleration factor	Scree
3	CNG	CNG
3	Optimal coordinates	Scree
3	Parallel analysis	Scree
3	Kaiser criterion	Scree
3	Scree (R2)	Scree_SE
4	beta	Multiple_regression
8	t	Multiple_regression
8	p	Multiple_regression
10	Scree (SE)	Scree_SE

11	Bentler	Bentler
13	Bartlett	Barlett
13	Anderson	Barlett
13	Lawley	Barlett

Otro método para determinar el número factores a extraer, es el destacado a partir del cálculo de los eigenvalues como resultado de la matriz de correlaciones, este método realiza una ponderación de acuerdo a las correlaciones registradas y estima según el peso de cada variable las mejores condiciones de las componentes que podrían explicar las variables entre sí, se presenta a continuación el resultado de su implementación:

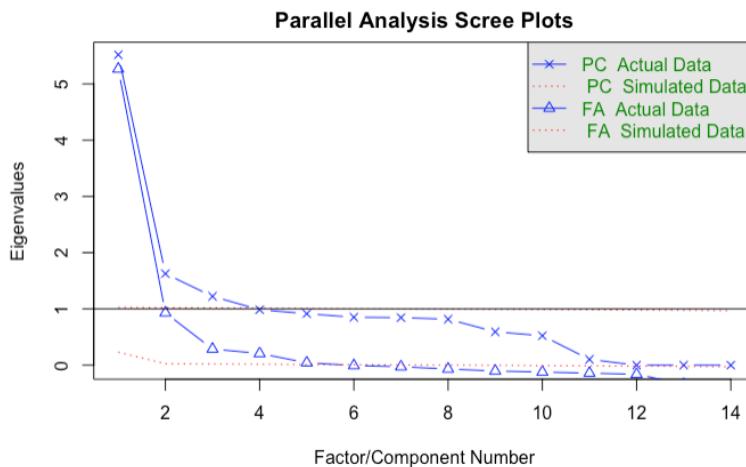
Gráfico 26. Gráfico método Scree AFE.



El diagrama de dispersión es una representación visual de los valores propios. La inspección visual del diagrama de dispersión es una forma rápida y sencilla de hacerse una idea de la dimensionalidad de su conjunto de datos, los valores que estén por encima de el peso de 1, son considerados óptimos, y las cantidades de componentes contados desde esa medición serían el número de factores a elegir. Es de notarse en el **Gráfico 26** el cuál muestra la aplicación del método scree, el 4to factor se encuentra más o menos en la "boundry"(límite) de la selección, resultando que el número de factores que deberían ser elegidos es de 3.

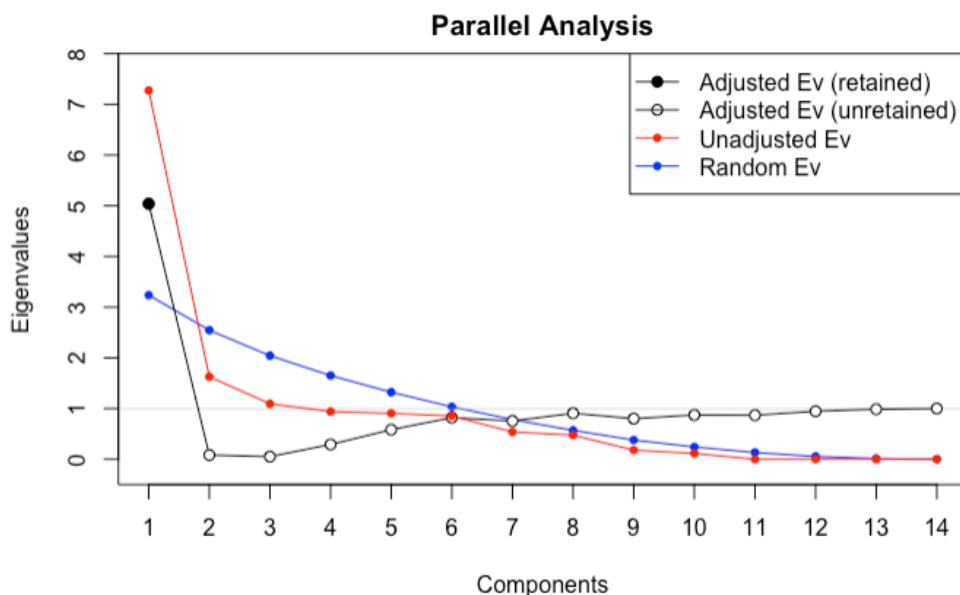
En la determinación del número de factores a extraer, se aplica el método de componentes principales iteradas (Principal Axes) de naturaleza no paramétrica, este se ocupa cuando no existe normalidad multivariante.

Gráfico 27. Gráfico Scree bajo el método Parallel.



En el método de Parallel presentado en el **Gráfico 27**, se determinó que el comportamiento de los componentes o factores a seleccionar es el mismo implementando la función metodológica de “PA(Principal Axes)”, y los componentes axiales de acuerdo a su simulación, condicionan una proyección no tan ajustada debido a que los valores del “Eigenvalues” son mejores explicados con los valores actuales.

Gráfico 28. Gráfico de análisis Parallel.



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Implementando los datos en diferentes instancias con análisis de sensibilidad de diferentes valores "Eigenvalues" tales como: El desajustado, el valor de "Eigenvalues" aleatorio y los ajustados, se obtiene la integración de gráficos en el **Gráfico 28**, los valores que están por encima para las diferentes variaciones estiman un promedio de 3,33 componentes obtenidos a partir de (Un componente para el "Adjusted EV", 3 componentes según el "Unadjusted EV" y 6 componentes de acuerdo al "Random EV", donde EV es "Eigenvalues").

Tabla 24. Resumen de cantidad de factores por cada procedimiento.

Consenso entre métodos: 3 factores	Scree plot: 3 factores	Parallel analysis: 3,33 factores
	Parallel analysis scree plot: 3 factores	

Las cantidades de factores que serían seleccionados conforme al resumen de todas las técnicas implementadas es de 3 (tres).

Una vez en la etapa de aplicación, se realizan diferentes métodos con el propósito de elegir el que mejores condiciones estadísticas permita realizar la selección de factores, implementando los algoritmos para cada método de procesamiento de datos se genera un consolidado de resultados a continuación.

Tabla 25. Comparación de métodos AFE aplicados al data set.

Tipificación	Métodos integrados	RMSR Adjust	Proporción de varianza acumulada explicada
Método 1: EFA1 "fa()" > data .	Cantidad de factores: 5, Tipo de rotación: Oblimin	0.03	60%
Método 2: EFA2 "factor.pa()" > data	Cantidad de factores: 5, Tipo de rotación: Varimax, Scores: TRUE	0.02	59%
Método 3: fa.mult.corr "fa.multi" > corr(data)	Cantidad de factores: 4, Tipo de rotación: Oblimin, Scores: Bartlett, Factores deseado de alto nivel: 2,	0.03	55%

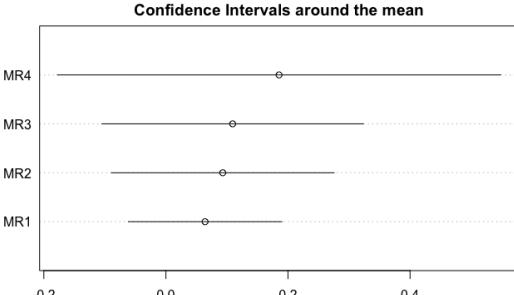
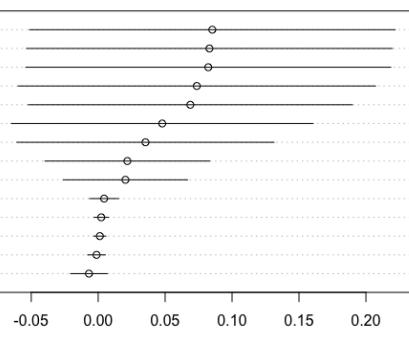
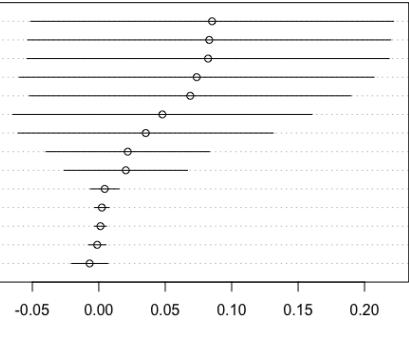
	Número de iteracciones bootstrap en fa: 1		
Método 4: EFA_model "fa()" >data.	Cantidad de factores: 3	0.04	51%
Método 5: datafactor "fa()" >data	Cantidad de factores: 3, Tipo de rotación: Varimax, Método de factorización: minres	0.04	51%
Método 6: cpi "fa()" >corr.test(data_stan) as.matrix(correlaciones\$r)	Cantidad de factores: 3, Tipo de rotación: Ninguna, Método de factorización: Ejes principales (pa)	0.04	54%

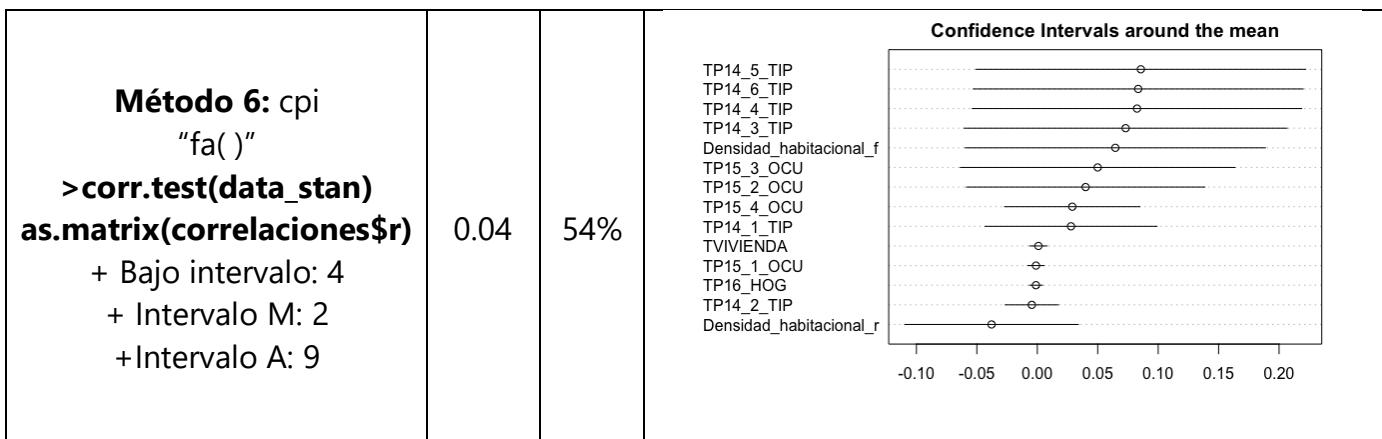
La aplicación de diversos métodos sin extensión de hiperparametros ajustados en función de los resultados del RMSR o la proporción de varianza explicada acumulada, se presentan en la **Tabla 25**, el mejor modelo es el que exprese resultados de varianza proporcional acumulada con mayor magnitud y el RMSR (Error cuadrático medio) más pequeño, con estas características es posible identificar que el método 1 cumple con la condición expresadas en comparación con los demás métodos y con el segundo mejor porcentaje de explicación de la varianza.

Estos indicadores estadísticos nos permiten tomar la decisión determinar que métodos es más estadísticamente significativo, por tanto a criterio de descarte se realizan los intervalos de confiabilidad de residuos para cada método de manera que se pueda apreciar la cobertura de error de estos métodos entendiéndose que entre menor sean los intervalos para todas las variables, mejor es el método, gráficamente se obtiene lo siguiente.

Tabla 26. Análisis de intervalos de confiabilidad para cada método AFE.

Tipificación	RMSR	PVAE	Intervalos de confiabilidad de la media de residuos																																																												
Método 1: EFA1 "fa()" >data + Bajo intervalo: 6 + Intervalo M: 3 +Intervalo A: 5	0.03	60%	<p style="text-align: center;">Confidence Intervals around the mean</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Mean (approx.)</th> <th>Lower Bound (approx.)</th> <th>Upper Bound (approx.)</th> </tr> </thead> <tbody> <tr><td>Densidad_habitacional_f</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_3_TIP</td><td>0.18</td><td>0.05</td><td>0.31</td></tr> <tr><td>TP14_6_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_4_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_5_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_3_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_2_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_4_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TVIVIENDA</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP16_HOG</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_2_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_1_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_1_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>Densidad_habitacional_r</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> </tbody> </table>	Variable	Mean (approx.)	Lower Bound (approx.)	Upper Bound (approx.)	Densidad_habitacional_f	0.02	-0.01	0.05	TP14_3_TIP	0.18	0.05	0.31	TP14_6_TIP	0.02	-0.01	0.05	TP14_4_TIP	0.02	-0.01	0.05	TP14_5_TIP	0.02	-0.01	0.05	TP15_3_OCU	0.02	-0.01	0.05	TP15_2_OCU	0.02	-0.01	0.05	TP15_4_OCU	0.02	-0.01	0.05	TVIVIENDA	0.02	-0.01	0.05	TP16_HOG	0.02	-0.01	0.05	TP14_2_TIP	0.02	-0.01	0.05	TP14_1_TIP	0.02	-0.01	0.05	TP15_1_OCU	0.02	-0.01	0.05	Densidad_habitacional_r	0.02	-0.01	0.05
Variable	Mean (approx.)	Lower Bound (approx.)	Upper Bound (approx.)																																																												
Densidad_habitacional_f	0.02	-0.01	0.05																																																												
TP14_3_TIP	0.18	0.05	0.31																																																												
TP14_6_TIP	0.02	-0.01	0.05																																																												
TP14_4_TIP	0.02	-0.01	0.05																																																												
TP14_5_TIP	0.02	-0.01	0.05																																																												
TP15_3_OCU	0.02	-0.01	0.05																																																												
TP15_2_OCU	0.02	-0.01	0.05																																																												
TP15_4_OCU	0.02	-0.01	0.05																																																												
TVIVIENDA	0.02	-0.01	0.05																																																												
TP16_HOG	0.02	-0.01	0.05																																																												
TP14_2_TIP	0.02	-0.01	0.05																																																												
TP14_1_TIP	0.02	-0.01	0.05																																																												
TP15_1_OCU	0.02	-0.01	0.05																																																												
Densidad_habitacional_r	0.02	-0.01	0.05																																																												
Método 2: EFA2 "factor.pa()" >data + Bajo intervalo: 3 + Intervalo M: 5 +Intervalo A: 5	0.02	59%	<p style="text-align: center;">Confidence Intervals around the mean</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Mean (approx.)</th> <th>Lower Bound (approx.)</th> <th>Upper Bound (approx.)</th> </tr> </thead> <tbody> <tr><td>Densidad_habitacional_f</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_3_TIP</td><td>0.18</td><td>0.05</td><td>0.31</td></tr> <tr><td>TP14_6_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_4_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_5_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_3_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_2_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_1_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_4_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TVIVIENDA</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP16_HOG</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_1_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_2_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>Densidad_habitacional_r</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> </tbody> </table>	Variable	Mean (approx.)	Lower Bound (approx.)	Upper Bound (approx.)	Densidad_habitacional_f	0.02	-0.01	0.05	TP14_3_TIP	0.18	0.05	0.31	TP14_6_TIP	0.02	-0.01	0.05	TP14_4_TIP	0.02	-0.01	0.05	TP14_5_TIP	0.02	-0.01	0.05	TP15_3_OCU	0.02	-0.01	0.05	TP15_2_OCU	0.02	-0.01	0.05	TP14_1_TIP	0.02	-0.01	0.05	TP15_4_OCU	0.02	-0.01	0.05	TVIVIENDA	0.02	-0.01	0.05	TP16_HOG	0.02	-0.01	0.05	TP15_1_OCU	0.02	-0.01	0.05	TP14_2_TIP	0.02	-0.01	0.05	Densidad_habitacional_r	0.02	-0.01	0.05
Variable	Mean (approx.)	Lower Bound (approx.)	Upper Bound (approx.)																																																												
Densidad_habitacional_f	0.02	-0.01	0.05																																																												
TP14_3_TIP	0.18	0.05	0.31																																																												
TP14_6_TIP	0.02	-0.01	0.05																																																												
TP14_4_TIP	0.02	-0.01	0.05																																																												
TP14_5_TIP	0.02	-0.01	0.05																																																												
TP15_3_OCU	0.02	-0.01	0.05																																																												
TP15_2_OCU	0.02	-0.01	0.05																																																												
TP14_1_TIP	0.02	-0.01	0.05																																																												
TP15_4_OCU	0.02	-0.01	0.05																																																												
TVIVIENDA	0.02	-0.01	0.05																																																												
TP16_HOG	0.02	-0.01	0.05																																																												
TP15_1_OCU	0.02	-0.01	0.05																																																												
TP14_2_TIP	0.02	-0.01	0.05																																																												
Densidad_habitacional_r	0.02	-0.01	0.05																																																												
Método 3: fa.mult.corr "fa.multi" >corr(data) + Bajo intervalo: 5 + Intervalo M: 2 +Intervalo A: 7	0.03	55%	<p style="text-align: center;">Confidence Intervals around the mean</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>Mean (approx.)</th> <th>Lower Bound (approx.)</th> <th>Upper Bound (approx.)</th> </tr> </thead> <tbody> <tr><td>Densidad_habitacional_f</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_6_TIP</td><td>0.18</td><td>0.05</td><td>0.31</td></tr> <tr><td>TP14_3_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_4_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_5_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_3_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_2_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_4_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_1_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TVIVIENDA</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP14_2_TIP</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP16_HOG</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>TP15_1_OCU</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> <tr><td>Densidad_habitacional_r</td><td>0.02</td><td>-0.01</td><td>0.05</td></tr> </tbody> </table>	Variable	Mean (approx.)	Lower Bound (approx.)	Upper Bound (approx.)	Densidad_habitacional_f	0.02	-0.01	0.05	TP14_6_TIP	0.18	0.05	0.31	TP14_3_TIP	0.02	-0.01	0.05	TP14_4_TIP	0.02	-0.01	0.05	TP14_5_TIP	0.02	-0.01	0.05	TP15_3_OCU	0.02	-0.01	0.05	TP15_2_OCU	0.02	-0.01	0.05	TP15_4_OCU	0.02	-0.01	0.05	TP14_1_TIP	0.02	-0.01	0.05	TVIVIENDA	0.02	-0.01	0.05	TP14_2_TIP	0.02	-0.01	0.05	TP16_HOG	0.02	-0.01	0.05	TP15_1_OCU	0.02	-0.01	0.05	Densidad_habitacional_r	0.02	-0.01	0.05
Variable	Mean (approx.)	Lower Bound (approx.)	Upper Bound (approx.)																																																												
Densidad_habitacional_f	0.02	-0.01	0.05																																																												
TP14_6_TIP	0.18	0.05	0.31																																																												
TP14_3_TIP	0.02	-0.01	0.05																																																												
TP14_4_TIP	0.02	-0.01	0.05																																																												
TP14_5_TIP	0.02	-0.01	0.05																																																												
TP15_3_OCU	0.02	-0.01	0.05																																																												
TP15_2_OCU	0.02	-0.01	0.05																																																												
TP15_4_OCU	0.02	-0.01	0.05																																																												
TP14_1_TIP	0.02	-0.01	0.05																																																												
TVIVIENDA	0.02	-0.01	0.05																																																												
TP14_2_TIP	0.02	-0.01	0.05																																																												
TP16_HOG	0.02	-0.01	0.05																																																												
TP15_1_OCU	0.02	-0.01	0.05																																																												
Densidad_habitacional_r	0.02	-0.01	0.05																																																												

			
Método 4: EFA_model "fa()" >data. + Bajo intervalo: 5 + Intervalo M: 2 +Intervalo A: 7	0.04	51%	
Método 5: datafactor "fa()" >data + Bajo intervalo: 5 + Intervalo M: 2 +Intervalo A: 7	0.04	51%	



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Los diagramas de intervalos de confiabilidad cerca de la media de residuos del cada modelo, siguen un patrón de rango de intervalo por variable, lo ideal es que todas las variables tengan las mediciones con centros de medias iguales a cero, es decir que los registros presentados ("círculos") representados para cada variable tiendan al valor más pequeño posible, estos reflejan la media de los residuos y las líneas los alcances que estas mediciones tienen tanto un límite superior como un límite inferior.

Evaluando tanto el rango del intervalo como la distribución de la media en cada variable, se realiza un conteo de las variables por tipo de intervalo, para el cual se tiene que menores a una medida de error por debajo de la escala de 0.05 se considera de "Bajo intervalo", el intervalo medio (Intervalo M) se configura para un rango valores a sus extremos mayores al 10% del error, y los intervalos amplios (Intervalo A) aquellos que superen el error de sus extremos del 10%.

De acuerdo a la selección del mejor modelo se tiene en cuenta aquellos que tengan la menor cantidad de variables con rango amplios y que cumplan con el supuesto de ajuste del análisis factorial. Teniendo en cuenta los análisis de límites de confiabilidad como la visualización de la variable del residuo en función de cada variables observada, se condiciona la decisión de selección del mejor modelo a partir de extensiones de pruebas estadísticas, respectivamente:

+Las variables que tienen el "violín chart" con área significativa se consideran variables con altos índices de error distribuido(AIED), las que tienen áreas muy pequeñas se consideran de bajo índice de error distribuido (BIED) y las que tienen áreas cuyos límites superiores e inferiores oscilen entre el 10% se consideran de índices medio de error distribuido (IMED).

Contando las cantidades de variables para cada categoría se obtiene el consolidado en la **Tabla 27**, en la cual se busca cuantificar los métodos que tengan mayor cantidad de variables con "BIED".

Tabla 27. Análisis de intervalos de confiabilidad de media residuo para cada método AFE.

Tipificación	RMSR	PVA E	Intervalos de confiabilidad de la media de residuos
Método 1: EFA1 "fa()" >data + BIED: 6 + IMED: 3 + AIED: 5	0.03	60%	
Método 2: EFA2 "factor.pa()" >data + BIED: 3 + IMED: 6 + AIED: 5	0.02	59%	
Método 3: fa.mult.corr "fa.multi" >corr(data) + BIED: 5 + IMED: 4 + AIED: 5	0.03	55%	

Método 4: EFA_model "fa()" >data. + BIED 5 + IMED 4 +AIED: 5	0.04 51%	
Método 5: datafactor "fa()" >data + BIED: 5 + IMED 4 +AIED 5	0.04 51%	
Método 6: cpi "fa()" >corr.test(data_stan) as.matrix(correlaciones\$r $)$ + BIED: 4 + IMED: 5 +AIED: 5	0.03 51%	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

La búsqueda e interpretación de los estadísticos de ajuste del modelo absoluto y relativo para seleccionar el mejor modelo que mejor se ajuste en los datos se realiza a partir del cumplimiento de los criterios estadísticos. Desde el procesamiento de datos, se planteó que las variables observadas precedían de un cargue teórico de tres factores. La confianza de selección no solo la dan un seriado de test que influyan en la determinación de factores a extraer, sino

en la utilización de los estadísticos de ajuste para tomar una decisión empírica sobre cuántos factores utilizar.

De acuerdo a lo anterior, las mejores opciones para elegir el mejor modelo son el “Método 1” para 5 factores y el “método 7” a partir de las condiciones de los indicadores creados que valide el menor valor de RMSR y la proporción de varianza acumulada explicativa, no obstante, por ultimo se aplicaría el test del “BIC” el cual es una estadística de estilo de ajuste relativo para comparar los modelos en términos de las cantidades de factores y determinar que modelo es el indicado para seleccionar.

Ilustración 41. Resultados del test BIC y Chi-Squared para cada modelo.

```
[1] "El valor del 'chi' del método 1 es:"
[1] 1773.42
[1] "El valor del 'chi' del método 2 es:"
[1] 1029.971
[1] "El valor del 'chi' del método 4 es:"
[1] 7836.699
[1] "El valor del 'chi' del método 5 es:"
[1] 7836.699
[1] "El valor del 'chi' del método 6 es:"
[1] 6942.248
[1] "El valor del 'BIC' del método 1 es:"
[1] 33276428066
[1] "El valor del 'BIC' del método 2 es:"
[1] 1.702712e+12
[1] "El valor del 'EBIC' del método 4 es:"
[1] 7281.423
[1] "El valor del 'BIC' del método 5 es:"
[1] 1.314331e+12
[1] "El valor del 'BIC' del método 6 es:"
[1] 1.627484e+12
```

En los resultados es posible apreciar que el valor de “chi-squared” de cada método es un estadístico confiable para determinar el mejor método de acuerdo a la forma en que las discrepancias relativas de los datos afectan la precisión de la explicación, en este caso la comparación se orienta para a los mínimos valores del test BIC, obteniendo que el “Método 4” es el mejor método para explicar los datos pero en el análisis de intervalos de la media de residuos indica que tiene el peor resultado en RMSA y el “Porcentaje de Varianza Acumulada Explicativa”(PVAE), por lo tanto es plausible solo considerar los métodos con mejor resultado en general, para estos perfiles el **método 1 y método 2**. El método 2, responde a una prueba de Chi-Squared un valor menor que el método 2, pero el resultado del método 1 es mejor en términos de BIC o, ponderando en gran medida el modelo con mejor representación y aceptación.

El panorama de los modelos da a conocer las mejores configuraciones para los parámetros de un modelo ajustado por características dadas por otros modelos analizados de origen optimo,

por ejemplo que el mejor método de factorización en los datos es "Ejes principales", porque permite garantizar el mantenimiento de la máxima varianza en un 60%, estos hallazgos podrían ser el insumo para entrar a validar de forma más extensa el modelo más óptimo. En el criterio de selección se opta por seleccionar el **Método 1** como el mejor de todos, debido a su bajo BIC y cumplimiento rotundo de todas las condiciones estadísticas.

Análisis factorial exploratorio para el Método 1:

En el proceso de explicación con los factores seleccionados, cuya la selección fue para el método 1, se muestran los resultados del AFE a continuación:

De acuerdo a las variables observadas en el resultado de la aplicación de EFA, se explican las siguientes métricas del análisis factorial exploratorio:

+MR_n: Cargas factoriales de cada variable.

+h²: Comunalidad(Varianza común explicada)

+u²: Especificidad (Varianza residual) o varianza no explicada.

Los valores se muestran el siguiente reporte de valores con cargas factoriales, Comunalidad y especificidad:

Tabla 28. Resultado de AFE método 1: Cargas factoriales.

	MR1 <S3: AsIs>	MR3 <S3: AsIs>	MR2 <S3: AsIs>	MR4 <S3: AsIs>	MR5 <S3: AsIs>	h2 <dbl>	u2 <dbl>	com <dbl>
TVIVIENDA	0.90	-0.10	0.10	0.04	0.02	1.00257074	-0.002570742	1.053111
TP14_1_TIP	0.03	-0.02	0.98	0.00	0.02	0.99559247	0.004407532	1.004543
TP14_2_TIP	0.95	-0.11	-0.25	0.04	-0.04	0.99452409	0.005475905	1.170726
TP14_3_TIP	0.21	0.09	-0.03	0.02	0.17	0.06912807	0.930871934	2.378431
TP14_4_TIP	0.05	-0.05	-0.05	-0.03	0.35	0.11516172	0.884838284	1.129889
TP14_5_TIP	0.02	-0.07	-0.08	0.00	0.41	0.15527583	0.844724171	1.124110
TP14_6_TIP	0.04	-0.06	-0.07	-0.05	0.35	0.11412671	0.885873290	1.205714
TP15_1_OCU	0.95	0.02	0.10	0.04	0.03	0.98559527	0.014404733	1.031310
TP15_2_OCU	0.40	-0.37	0.06	-0.19	-0.05	0.39255474	0.607445260	2.539396
TP15_3_OCU	0.07	-0.04	0.01	0.79	0.00	0.70049987	0.299500133	1.022619
TP15_4_OCU	0.33	-0.63	0.01	-0.06	-0.01	0.69330433	0.306695672	1.537586
TP16_HOG	0.99	0.09	0.11	0.03	0.02	1.00168540	-0.001685402	1.041472
Densidad_habitacional_f	0.33	0.38	0.05	-0.07	0.01	0.14014193	0.859858073	2.079799
Densidad_habitacional_r	0.04	1.02	-0.02	-0.07	0.00	1.03873868	-0.038738677	1.012040

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

** A mayor h², mejor será el modelo 0;1

** A menor u², mejor será el modelo 0;1

Conclusión:

La variable "TVIVIENDA" es explicada en un 100%, "TP14_1_TIP" es explicada en un 99,55%, "Densidad_habitacional_r" es explicada al 100%, como se puede apreciar en la **Tabla 28** y la variable "Densidad_habitacional_f" solo es explicada en 14%, esto es que se pierde información de acuerdo a su composición de los datos.

Ilustración 42. Reporte de resultados del AFE: Método 1.

```

Factor Analysis using method = minres
Call: fa(r = data, nfactors = 5, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix

          MR1   MR3   MR2   MR4   MR5
SS loadings  4.24  1.90  1.10  0.72  0.44
Proportion Var 0.30  0.14  0.08  0.05  0.03
Cumulative Var 0.30  0.44  0.52  0.57  0.60
Proportion Explained 0.50  0.23  0.13  0.09  0.05
Cumulative Proportion 0.50  0.73  0.86  0.95  1.00
  
```

```

With factor correlations of
      MR1   MR3   MR2   MR4   MR5
MR1  1.00 -0.52  0.29  0.42  0.08
MR3 -0.52  1.00  0.05 -0.32  0.11
MR2  0.29  0.05  1.00  0.00  0.35
MR4  0.42 -0.32  0.00  1.00  0.06
MR5  0.08  0.11  0.35  0.06  1.00
  
```

The root mean square of the residuals (RMSR) is 0.01
The df corrected root mean square of the residuals is 0.03

The harmonic number of observations is 43408 with the empirical chi square 1773.42 with prob < 0

The total number of observations was 43408 with Likelihood Chi Square = 33276428397 with prob < 0

Tucker Lewis Index of factoring reliability = -31807.57
RMSEA index = 157.254 and the 90 % confidence intervals are 157.252 NA
BIC = 33276428066
Fit based upon off diagonal values = 1

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

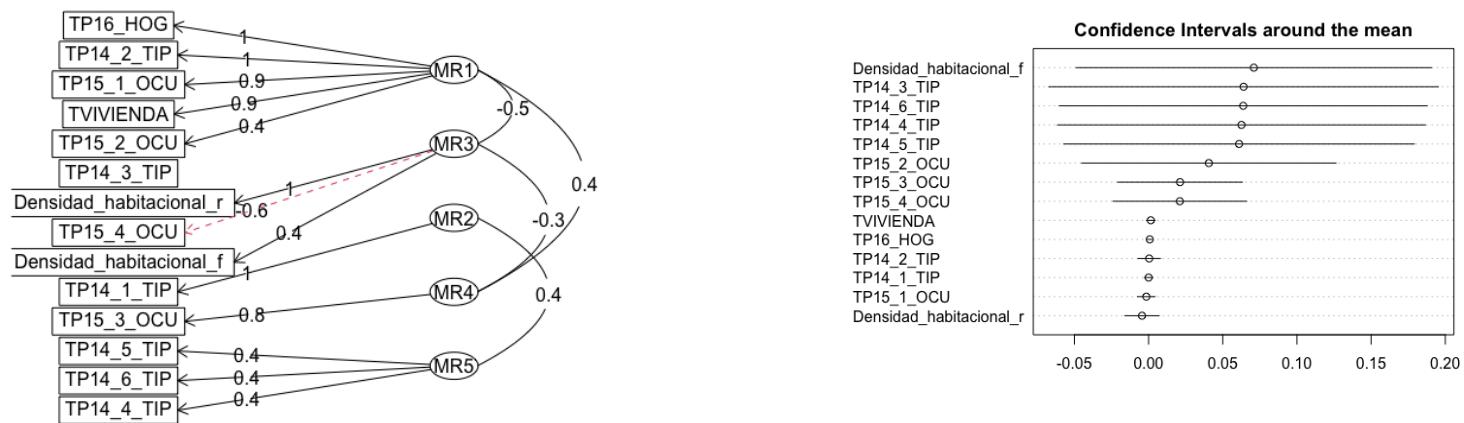
**Lo ideal es que la "Proportion Var" acumulada sea lo más cercano a 1.

+The corrected root mean square of the residuals (RMSR) is 0.03, correspondiente a la raíz cuadrada media de los residuos, este corresponde a un valor teórico de los residuos del modelo AFE. Es un buen indicador cuando su valor es menor que 0.08, en este caso presenta excelente resultado y la varianza explicada acumulada para todos los factores es de 60% por tanto es posible generar planteamientos a partir de ello resaltando que el método 1 es de mayor aceptación.

Los resultados obtenidos del AFE se explican a partir de los gráficos y ecuaciones factoriales. A continuación se presentan los diagramas explicativos del análisis factorial para el método seleccionado.

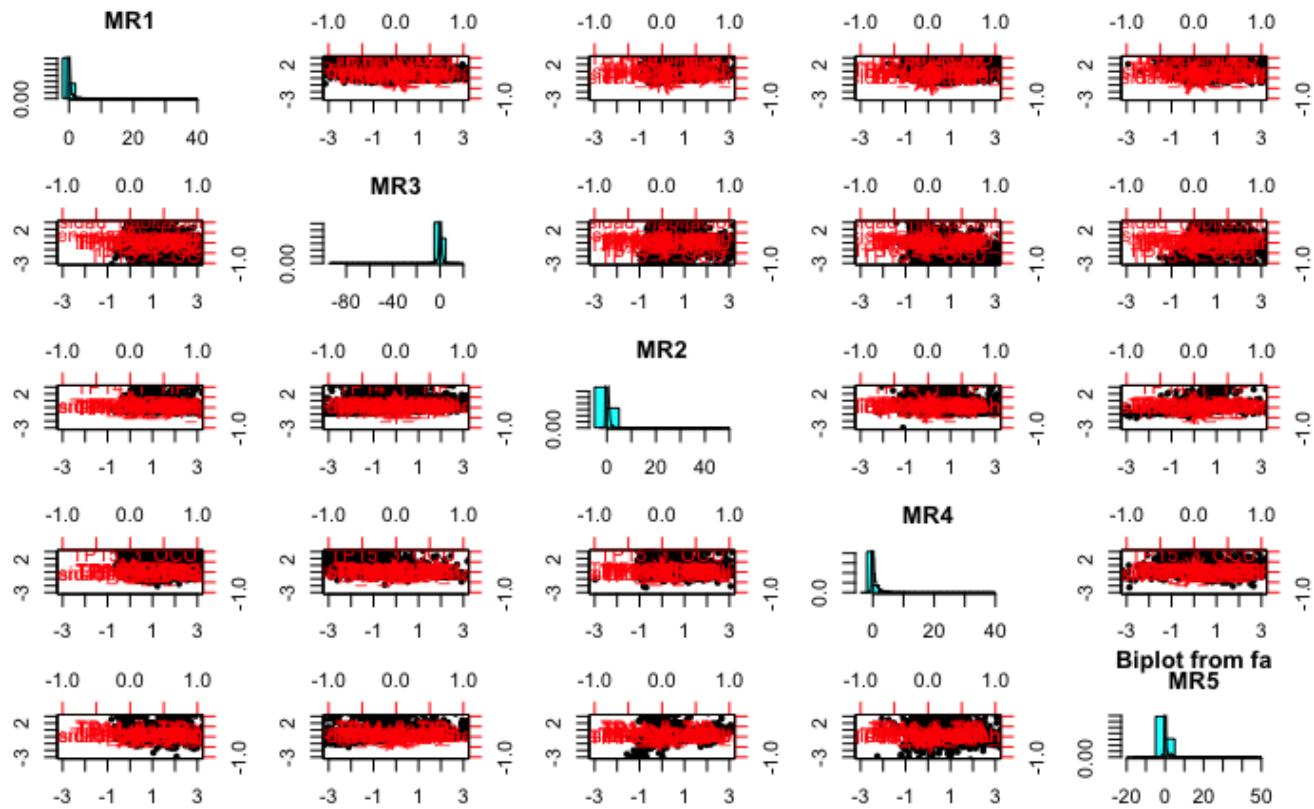
Diagrama 8. Análisis factorial exploratorio gráfico.

Análisis factorial confirmatorio para el indicador Densidad Habitacional M1



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

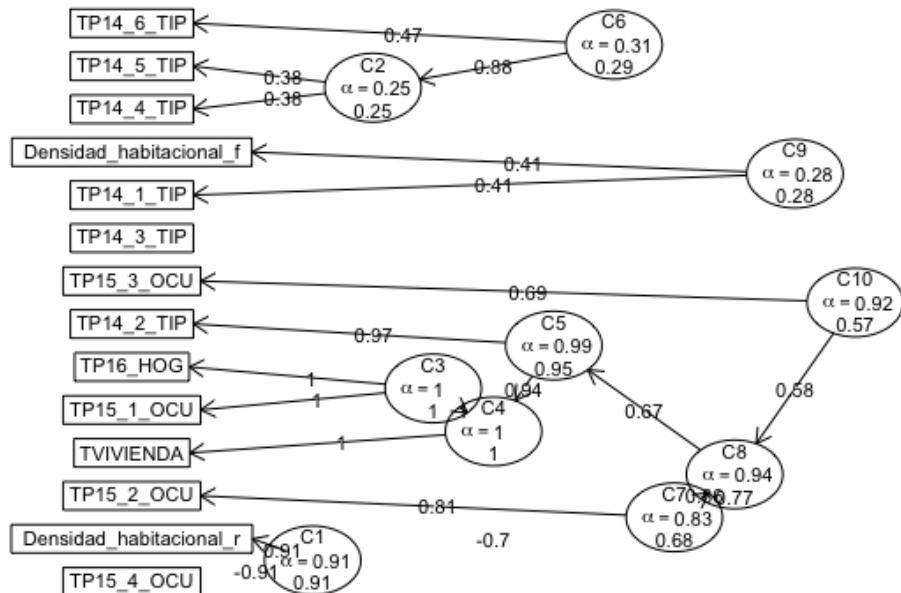
En este primer **Diagrama 8** se aprecian los factores seleccionados y la relación ponderada explicativa de estos factores con respecto a las variables observadas, es de observarse que para el Tercer factor (MR3) las variables de aproximación al indicador del "Densidad habitacional resta" es explicada al 100% y el indicador cuya aproximación proviene de una operación de resta de las variables de conteo de hogar y vivienda solo el 0.4, estas puntuaciones estiman el valor explicativo de ese factor con respecto a la variable "Densidad habitacional fracción". En ese orden de idea, los factores presentados al modelo AFE, consideran en su ajuste relativo conforme a las condiciones estadísticas de cada variable y establece las métricas para las variables en un orden de importancia, por tanto la variable más importante para el factor MR3 es "Densidad_habitacional_r". En continuidad al contexto, podría decirse que para el factor MR2 la única variable que refiere de mayor importancia es la "TP14_1_TIP" es decir la variable conteo de viviendas de tipo casa.

Gráfico 29. Matriz de dispersión entre factores.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

En el **Gráfico 29**, se muestra el despliegue de gráficos de dispersión de los datos, para el cual los valores representados son las concentraciones de los valores de correlación entre factores. Bajo la extensión de un algoritmo capaz de realizar agrupaciones de variables para explicarlas de forma que el análisis de fiabilidad de las variables observadas permita establecer las relaciones entre este grupo en función y su relacionamiento conjunto, se muestra la aplicación del método descrito "iClust", que realiza las agrupaciones propiamente dichas.

Diagrama 9. Diagrama ICLUS indicator Densidad Habitacional



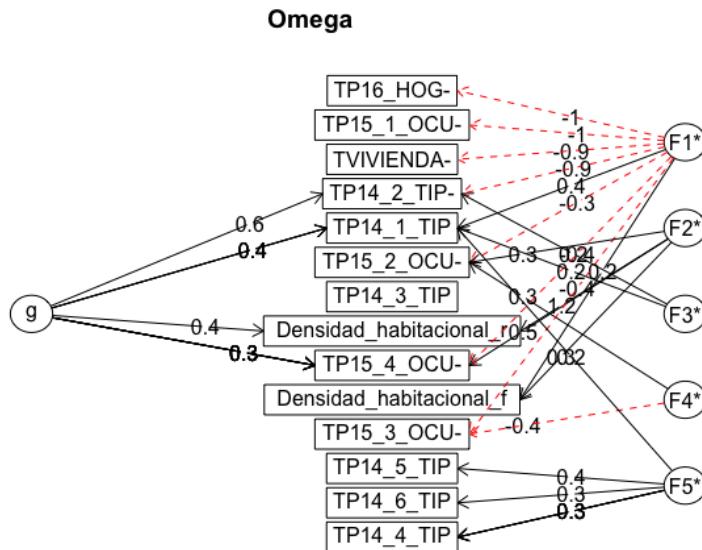
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Obteniendo un total de 10 clústeres, cuyos resultados se afirma que las aproximaciones de indicadores están contenidas en diferentes clústeres, en el clúster 1 (C1) guarda un fuerte relación explicativa con el "Densidad_habitacional_r" y la puntuación de fiabilidad es de 0.91, un resultado favorable que expresa que existe un excelente consistencia interna entre los grupos de ese clúster, superando la aceptabilidad del margen de 0.7; y para el clúster 9 (C9) de "Densidad_habitacional_f" tiene una puntuación explicativa y de importancia de 0.41, sus resultados de fiabilidad son 0.28, por tanto no son muy positivos.

(UCLA, s.f.)

El modelo bifactorial de omega de los datos, es una opción valida transicional a la hora de considerar el análisis factorial confirmatorio y es una alternativa al coeficiente alpha, por tanto la medida del omega garantiza que el modelo no se encuentre saturado, y permite entender la distribución de las cargas de fuerza de asociación entre las variables.

Diagrama 10. Modelo bifactorial Omega



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

El modelo bifactorial generalmente influye en todas las variables observadas y los factores específicos, estos capturan la covariación entre subconjuntos de ítems que permanecen por encima de la covarianza debida al factor general, incluyendo un factor de coste general que se relaciona en todos los ítems junto con cinco factores específicos que capturan el exceso de covarianza entre los ítems del mismo dominio de contenido. El **Diagrama 10**, presentado hace alusión a la ruta de este modelo, en el que se observa que cada ítem tiene una carga de factor general no nula (es decir, el factor general “g” que influye en todos las variables observadas) junto con una carga no nula en un factor específico perteneciente al dominio de contenido del ítem (es decir, cada factor específico “Fi” influye sólo en un subconjunto de variable observada). Dado que cada variable observada está directamente influenciado por dos factores o más factores, la ecuación de este modelo es una ecuación de regresión múltiple en la que cada ítem se reajusta simultáneamente en el factor general y en uno de los factores específicos. En teoría, el factor general debe estar des correlacionado con los factores específicos para garantizar la identificación del modelo, mientras que en otros modelos de AFC, todos los factores se correlacionan libremente entre sí (permitir que el factor general en un modelo bifactorial se correlacione con uno de los factores específicos provoca errores como la no convergencia o soluciones inadecuadas).

(SAGE JOURNALS, 2020)

Ilustración 43. Calculo de las cargas factoriales por cada variable (Agrupación en términos de peso).

Loadings:	MR1	MR3	MR2	MR4	MR5
TVIVIENDA	0.898		0.101		
TP14_1_TIP			0.981		
TP14_2_TIP	0.953	-0.109	-0.248		
TP14_3_TIP	0.213				0.171
TP14_4_TIP					0.351
TP14_5_TIP					0.413
TP14_6_TIP					0.354
TP15_1_OCU	0.946		0.105		
TP15_2_OCU	0.396	-0.374		-0.195	
TP15_3_OCU					0.788
TP15_4_OCU	0.333	-0.631			
TP16_HOG	0.989		0.108		
Densidad_habitacional_f	0.328	0.377			
Densidad_habitacional_r		1.015			

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

En las cargas de cada variable como resultado del análisis factorial exploratorio representada en la **Ilustración 43**, se contempla que la variable "Densidad_habitacional_f", debe explicarse por dos factores y en valores muy bajo de puntuación factorial, en cambio la variable "Densidad_habitacional_r", es la que mejor se relaciona con el factor 3 (MR3).

Análisis alpha de cronbach:

Aplicando el test "alpha" para mirar el grado de fiabilidad de las variables a fines al índice DH, se obtiene:

Ilustración 44. Análisis de Alpha de Cronbach: Intervalo de confianza.

```
Reliability analysis
Call: alpha(x = data)

lower alpha upper      95% confidence boundaries
0.81 0.81 0.81

Reliability if an item is dropped:

Item statistics
```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Con un 95% de confianza es posible afirmar que el alfa de cronbach de los datos es equivalente a 0.81, este alfa se considera muy bueno, no obstante, se busca mejorarlo a partir de la eliminación de variables que no aportan a la fiabilidad del conjunto de variables observadas.

Tabla 29. Resultado del Alpha de Cronbach aproximaciones normales.

raw_alpha <dbl>	std.alpha <dbl>	G6(smc) <dbl>	average_r <dbl>	S/N <dbl>	ase <dbl>	mean <dbl>	sd <dbl>	median_r <dbl>	
		raw_alpha <dbl>	std.alpha <dbl>	G6(smc) <dbl>	average_r <dbl>	S/N <dbl>	alpha se <dbl>	var.r <dbl>	med.r <dbl>
0.8093799	0.7252615	0.7381957	0.158645	2.639825	0.0002577476	16.61406	28.47899	0.08282071	
TVIVIENDA	0.7266644	0.6434353	0.6804291	0.1218910	1.804540	0.0004676307	0.09209662	0.07582773	
TP14_1_TIP	0.8036202	0.7077778	0.7578604	0.1570513	2.422054	0.0001340956	0.12970451	0.07605977	
TP14_2_TIP	0.7474428	0.6643844	0.7039811	0.1321530	1.979599	0.0003524258	0.09491218	0.08218443	
TP14_3_TIP	0.8126297	0.7260762	0.8633200	0.1693636	2.650651	0.0002555335	0.13195052	0.07978831	
TP14_4_TIP	0.8141893	0.7341006	0.9712712	0.1751699	2.760822	0.0002592677	0.13126317	0.08364149	
TP14_5_TIP	0.8141900	0.7341803	0.9705799	0.1752289	2.761948	0.0002592674	0.13123430	0.08769018	
TP14_6_TIP	0.8141410	0.7355282	0.9263150	0.1762309	2.781121	0.0002591353	0.13104522	0.08769018	
TP15_1_OCU	0.7265483	0.6440130	0.6823519	0.1221609	1.809092	0.0004374997	0.09522828	0.07514695	
TP15_2_OCU	0.8115311	0.7044208	0.8650354	0.1549216	2.383188	0.0002645434	0.11814589	0.08218443	
TP15_3_OCU	0.8125673	0.7116802	0.8797044	0.1595754	2.468371	0.0002634724	0.12381447	0.08291535	
TP15_4_OCU	0.8042926	0.6968316	0.8258891	0.1502433	2.298496	0.0002868345	0.10623424	0.08218443	
TP16_HOG	0.7263899	0.6419841	0.6797680	0.1212162	1.793172	0.0004412995	0.09643749	0.07542178	
Densidad_habitacional_f	0.8141309	0.7317262	0.7350924	0.1734242	2.727536	0.0002587930	0.13062261	0.08291535	
Densidad_habitacional_r	0.8283515	0.7974026	0.8653210	0.2323997	3.935897	0.0002204702	0.07809799	0.09124228	

**raw_alpha > 0.7 -> aceptable ; raw_alpha > 0.8 -> bueno ; raw_alpha > 0.9 -> excelente.

En los raw_alpha es posible apreciar que ninguna variable tiene magnitudes negativas, por lo que se podría inferir que todas las variables contribuyen de forma positiva al incremento o sostenimiento del alpha de cronbach.

Se procede entonces a rescalar los valores del indicador para cada aproximación, con la finalidad de determinar el de mayor influencia de alpha de cronbach para el conjunto de datos, respectivamente:

Tabla 30. Resultado del Alpha de Cronbach aproximaciones rescaladas.

	raw_alpha <small><dbl></small>	std.alpha <small><dbl></small>	G6(smc) <small><dbl></small>	average_r <small><dbl></small>	S/N <small><dbl></small>	alpha se <small><dbl></small>	var.r <small><dbl></small>
TVIVIENDA	0.7510715	0.6434353	0.6804291	0.1218910	1.804540	0.0003969888	0.09209662
TP14_1_TIP	0.8192615	0.7077778	0.7578604	0.1570513	2.422054	0.0001151307	0.12970451
TP14_2_TIP	0.7680252	0.6643844	0.7039811	0.1321530	1.979599	0.0002776084	0.09491218
TP14_3_TIP	0.8268915	0.7260762	0.8633200	0.1693636	2.650651	0.0002175067	0.13195052
TP14_4_TIP	0.8282969	0.7341006	0.9712712	0.1751699	2.760822	0.0002203734	0.13126317
TP14_5_TIP	0.8282975	0.7341803	0.9705799	0.1752289	2.761948	0.0002203697	0.13123430
TP14_6_TIP	0.8282525	0.7355282	0.9263150	0.1762309	2.781121	0.0002202348	0.13104522
TP15_1_OCU	0.7509519	0.6440130	0.6823519	0.1221609	1.809092	0.0003648281	0.09522828
TP15_2_OCU	0.8257780	0.7044208	0.8650354	0.1549216	2.383188	0.0002221128	0.11814589
TP15_3_OCU	0.8267699	0.7116802	0.8797044	0.1595754	2.468371	0.0002229968	0.12381447
TP15_4_OCU	0.8187819	0.6968316	0.8258891	0.1502433	2.298496	0.0002286339	0.10623424
TP16_HOG	0.7516471	0.6419841	0.6797680	0.1212162	1.793172	0.0003657154	0.09643749
Densidad_habitacional_f.rescalado	0.8282915	0.7317262	0.7350924	0.1734242	2.727536	0.0002203612	0.13062261
Densidad_habitacional_r.rescalado	0.8283105	0.7974026	0.8653210	0.2323997	3.935897	0.0002203805	0.07809799

Conclusión : Aplicando el rescalado de valores a las aproximaciones, los valores del raw_alpha para la aproximaciones resultan ser muy próximas entre ellas, es decir que cualesquiera de estas que se elimine el nivel con que se consolida la fiabilidad de la información en todas las variables será igual. De acuerdo a todas las premisas fundamentales expuestas anteriormente, con los soportes estadísticos suficientes para garantizar una adecuada explicación del fenómeno de la densidad habitacional, se concluye que la operación “Resta” (Conteo de Hogares – Conteo de viviendas) resulta ser estadísticamente significativo a la hora de explicar el conjunto de 12 variables en las que se discriminan y categorizan los tipos de viviendas establecido a partir del estudio del MGN-NIM-2018, por lo tanto se logra concluir que el indicador de relacionamiento de hogares y la cantidad de viviendas está dado por la variable “Densidad habitacional por operación resta”.

Análisis categórico exploratorio del indicador: “Densidad habitacional”.

En la creación de indicadores, se deben tener en cuenta que la información puede tener mucha variación por influencia de las variables confusoras, es decir existe una influencia directa sobre los resultados de la medición de un indicador para un conjunto de poblaciones, por tanto en este capítulo se busca mitigar esa “perturbación” que se obtiene a nivel de localidad, considerando la “estandarización directa” de los datos.

Tipificando el indicador se obtiene el siguiente diagrama de densidad de los datos:

Gráfico 30. Gráfico de densidad del indicador DH por localidad**MGN- Nivel Manzana :Dataset 2021**

El diagrama de densidad del indicador al ser tipificado muestra que se presentan valores muy por debajo de los promedios ideales, es decir que la cantidad de hogares, sean iguales a la cantidad de viviendas (cero), por el contrario se muestran una alta presencia de datos atípicos. Se clasifican las categorías con los siguientes rangos de valores:

- $[-\infty, -90] \rightarrow \text{"Alto Déficit"}$
- $[-90, -6] \rightarrow \text{"Moderado Déficit"}$
- $[-6, -0.001] \rightarrow \text{"Déficit"}$
- $[-0.001, 0.001] \rightarrow \text{"Equilibrio"}$
- $[0.001, 6] \rightarrow \text{"Superavit"}$
- $[6, 13] \rightarrow \text{"Alto Superavit"}$

A continuación, se establece el resumen estadístico de estas categorías:

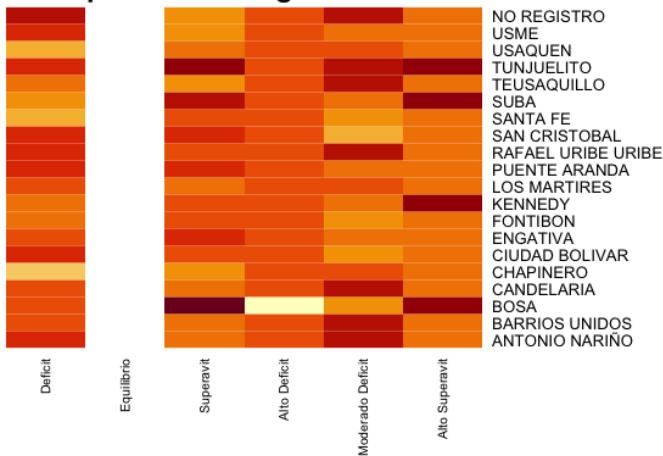
Ilustración 45. Resumen de categorías indicador Densidad Habitacional

Alto Deficit	Moderado Deficit	Deficit	Equilibrio	Superavit	Alto Superavit
1	109	16503	13230	13544	21

Como es de apreciarse en la **Ilustración 45**, que el 30,47% de los datos se concentran en la categoría "Equilibrio" es decir que el más del 30% de las manzanas en general cuenta con una relación 1 a 1 de hogares, las cantidades de hogares son iguales o equivalentes a las cantidades de viviendas.

Ilustración 46. Mapa de calor del indicador Densidad Habitacional

Heatmap tabla contingencia indicador

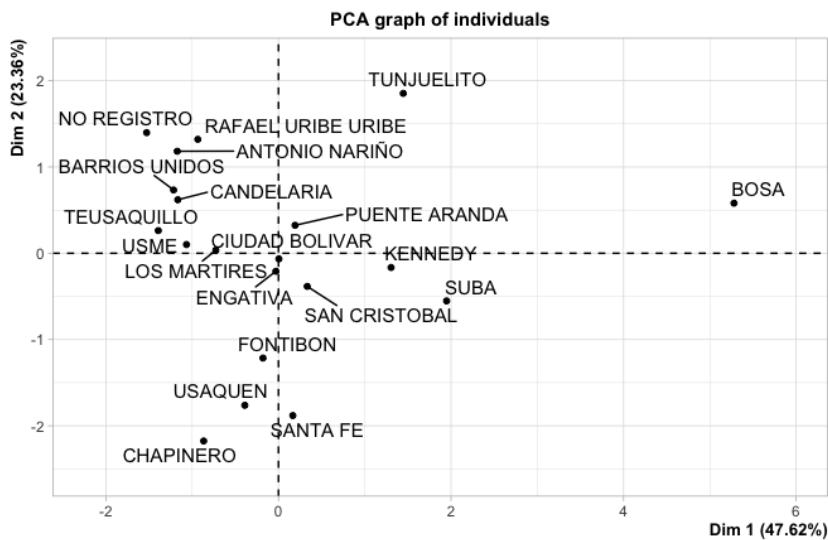


En la localidad de "Bosa", se presenta que en promedio las cantidades de manzanas que coinciden con el indicador en las categorías establecidas presentan mayores niveles del indicador comparado con las demás localidades, esto quiere decir que el indicador tienen mayor presencia en esta localidad en el mismo nivel que "Kennedy", "Suba" y "Tunjuelito", esto indica que existe un alto superávit de hogares más específicamente se evidencia una elevada proporción de hogares mayores que las

cantidades de viviendas en promedio en esas localidades. Así mismo en la localidad de Bosa, se presenta que existe una gran proporción de viviendas que son mayores en términos de cantidad que la población de hogares, la localidad de Teusaquillo muestra un alto valor de densidad habitacional con un déficit de hogares, esto indica que las cantidades de viviendas situadas en la localidad es mayores que las cantidades de hogares.

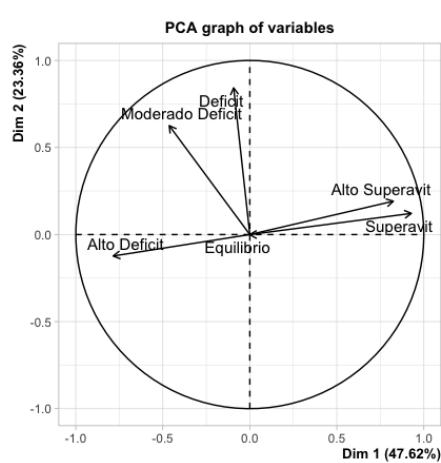
Aplicando un análisis de componentes principales exploratorio, se evidencia que dos componentes explican 70,98% de la información de los datos procesados, respectivamente: El método espacial individual se muestra en el **Diagrama 11**.

Diagrama 11.PCA Método Gráfico Espacial : índice DH



Las localidades que representan el peso de participación de cada localidad en función del indicador del déficit absoluto habitacional se encuentran muy próximas unas a otras, esto es que tienen perfiles muy similares a diferencia de la localidad de "Bosa", quien esta alejada de las localidades debido a condiciones especiales marcadas por el indicador, es decir esta tiene un perfil en el que registrada alto estándares de superávit de hogares y déficit.

Diagrama 12. PCA método circular.



Es evidente que las categoría "Alto déficit" y "Alto superávit" van en diferentes direcciones, no obstante lo que respecta a las categorías "Moderado Déficit" y "Déficit", están en el mismo cuadrante lo que equivale a decir que existe un alta correlación con estas dos entidades.

Para el 70,98% de la varianza de los datos, las dos componentes explican que la mayoría de localidades de acuerdo con el indicador estarían ubicadas en dirección de "Equilibrio". De las 20 localidades, 18 tienen los mismo puntos fuertes y los mismos puntos débiles de los valores de indicador y como todas estas están ubicadas muy próximas al origen indican que presentan distribuciones de valores para el valor esperado del indicador, es decir estas no se diferencian unas con otras en ningún atributo. En primer cuadrante la escala se expresa las categorías de "Superávit" y "Alto Superávit" y una dispersión significativa de la localidad "Bosa".

para el valor esperado del indicador, es decir estas no se diferencian unas con otras en ningún atributo. En primer cuadrante la escala se expresa las categorías de "Superávit" y "Alto Superávit" y una dispersión significativa de la localidad "Bosa".

Gráfico 31. Gráfico de tablas cruzadas del indicador Densidad Habitacional por categorías.

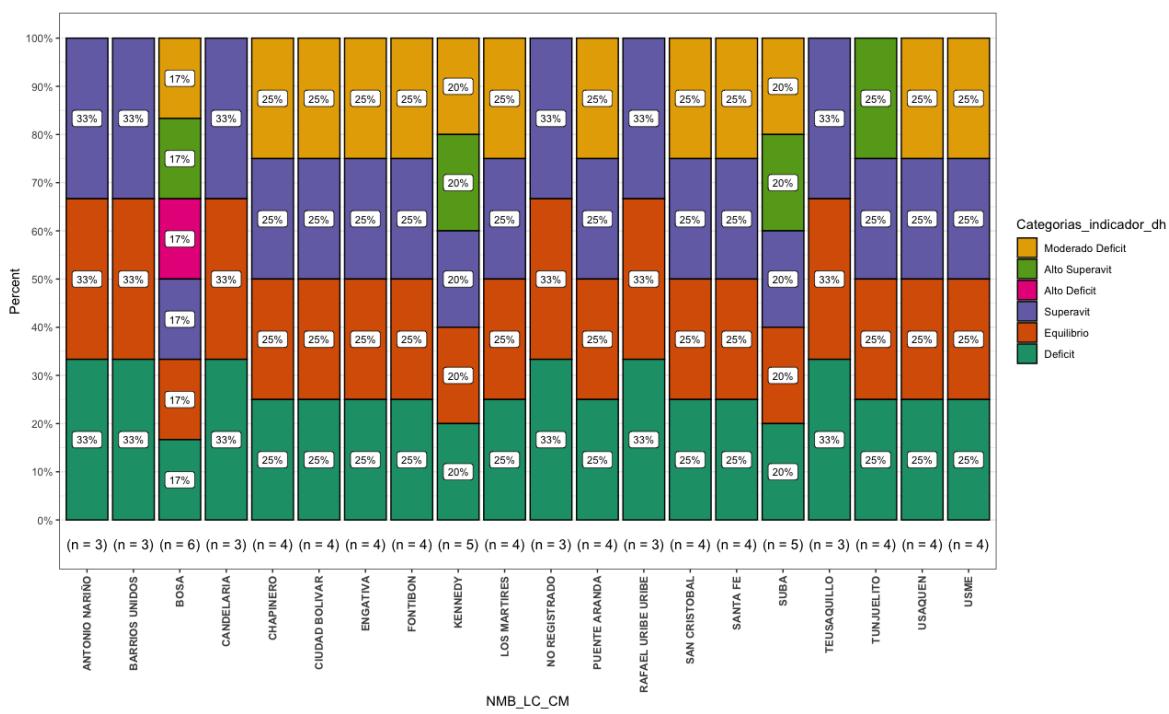
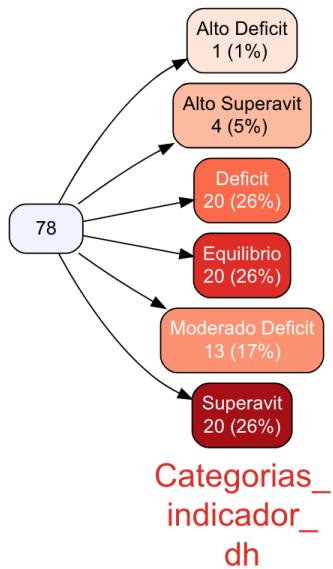


Diagrama 13. Diagramas de tablas cruzadas del indicador Densidad Habitacional por categorías.



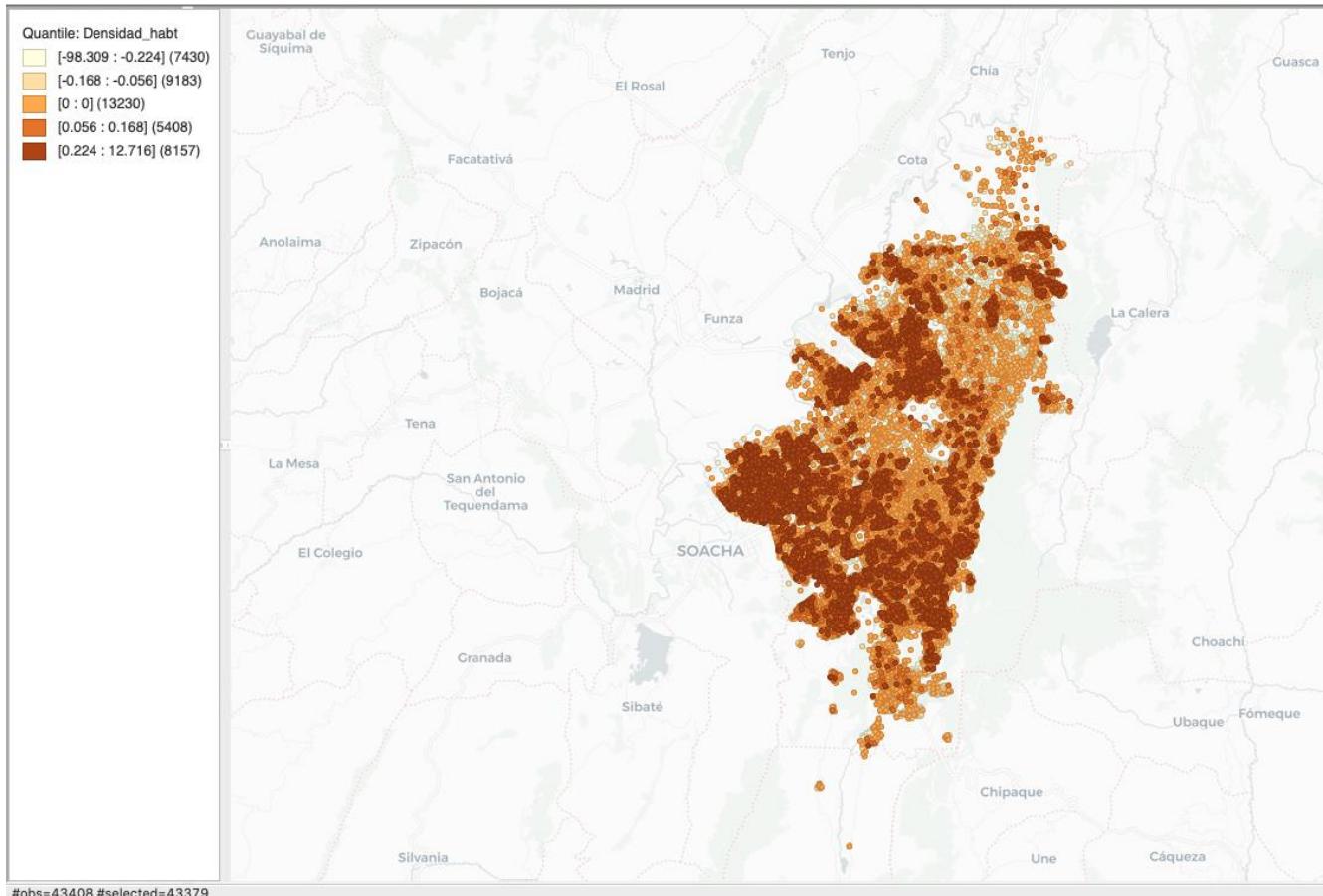
En el **Gráfico 31**, se encontró que 10 localidades cuentan con igualdad de participación para cuatro categorías tales como : "Superávit", "Déficit", "Equilibrio" y "Moderado Déficit", estableciendo un rango de participación en un intervalo de valores de [-90,6], estas refieren que las localidades en sus cantidades de manzanas reflejan un intervalo de confianza en función de la proporción de las categorías mencionadas, por tanto es normal que dentro de estas localidades 1 de cada 4 manzanas se encuentren en cualquiera de estas categorías mencionadas. Las localidades "Rafael Uribe Uribe", "Teusaquillo", "Antonio Nariño", "Barrios Unidos" y "Candelaria" presentan que 1 de cada 3 manzanas se encuentran en "Déficit", "Equilibrio" y "Superávit", es de señalarse así mismo que para las localidades no registradas; en "Kennedy" y "Suba" 1 de cada 5 manzanas se encuentran en todas las categorías a excepción del "Alto déficit". Finalmente, es de reconocer que la localidad "Bosa", refleja condiciones únicas del índice de densidad habitacional presentando participación en todas las categorías, es decir que 1 de cada 6 manzanas se encuentran en cualquiera de las categorías.

Se calculan 78 manzanas representativas en el **Diagrama 13**, que agrupan los promedios para cada categoría, las cuales devienen diferentes participaciones, correspondientes a:

Las categorías "Déficit", "Equilibrio", "Superávit" con 26% cada una, "Moderado Déficit" con un 17%, acumulando un total del 94% de las manzanas representativas, el restante se agrupa para las manzanas representativas con un 4% para la categoría "Alto Superávit" y 1% "Alto déficit".

Análisis espacial de los resultado de la medición del indicador “Densidad habitacional”:
La clasificación del indicador por cuartiles, se presenta a continuación :

Ilustración 47. Mapa de distribución geoespacial del indicador DH en Cuartiles (5)



La clasificación del indicador por cuartiles muestra la participación por nivel del indicador, presentando que número máximo (9183) manzanas de Bogotá en la clasificación, se encuentran entre los rangos de [-0.168 , -0.056] es decir entre una categoría de tipo “Déficit de Hogares”.

Una clasificación que contemple datos atípicos, conforme a 1.5 veces el rango intercuartílico de la distribución de los datos, y que muestra así la presencia de Outliers, de máximos y mínimos se obtiene en la siguiente visualización:

Ilustración 48. Mapas de distribución geoespacial del indicador DH en Cuartiles (1.5 veces el Rango Intercuartílico). (Gráfico de clasificación intercuartílico)

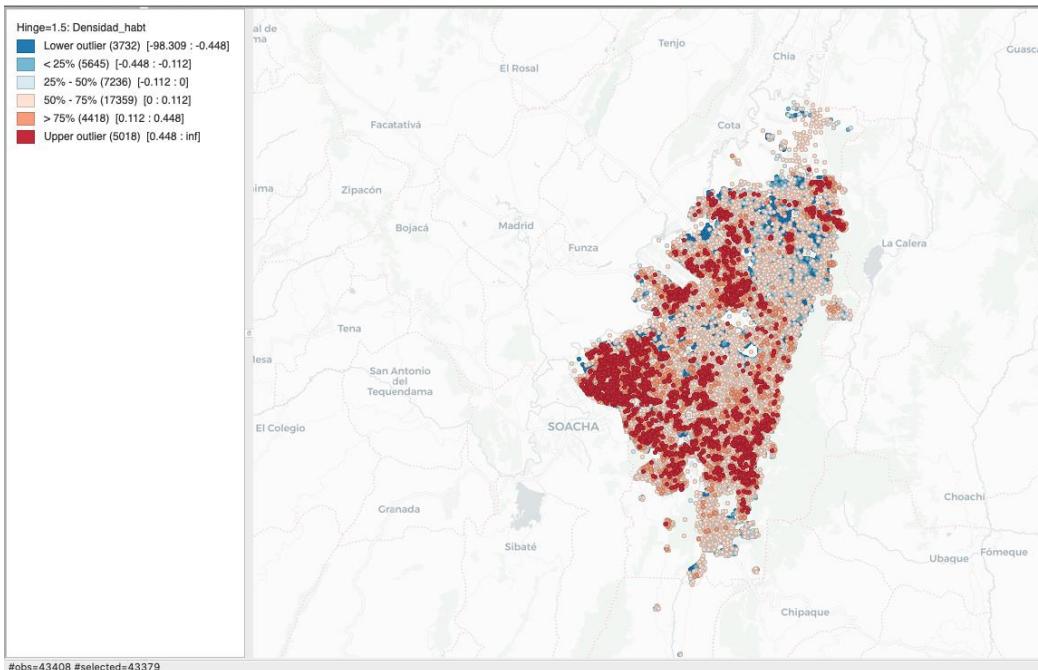
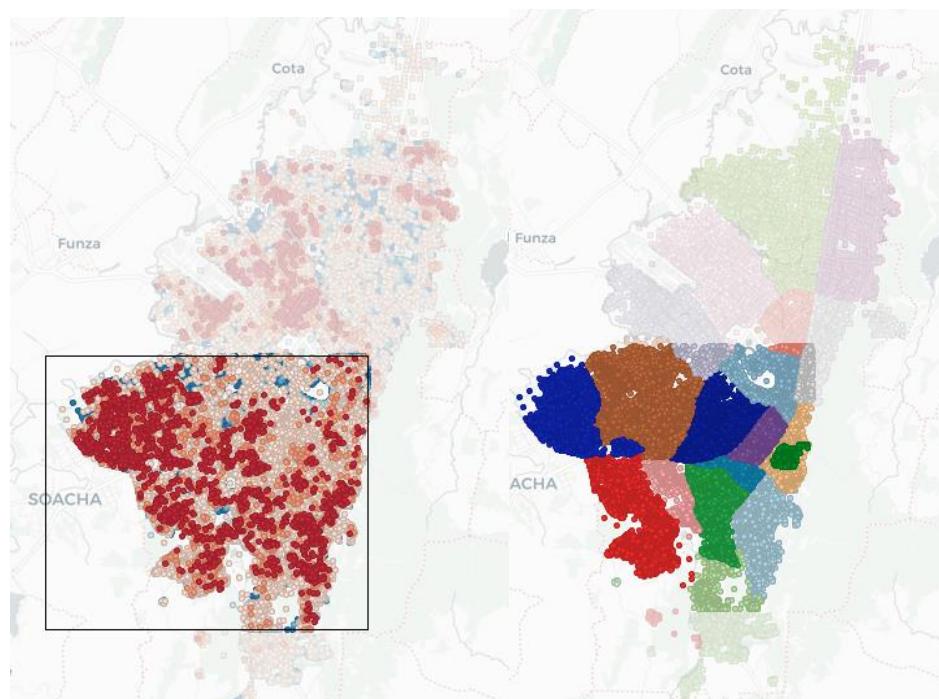
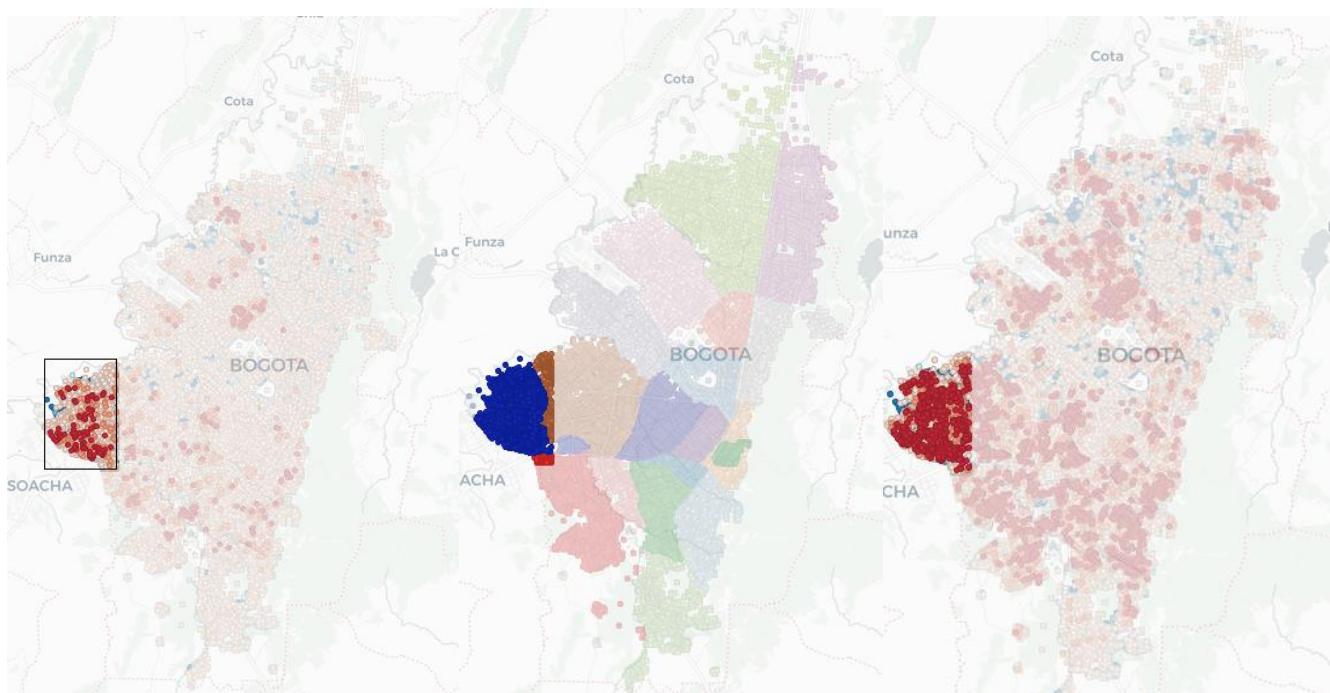


Ilustración 49. Mapas segmentados de comparación de poblaciones con valores en Alto Superávit de Densidad Habitacional.



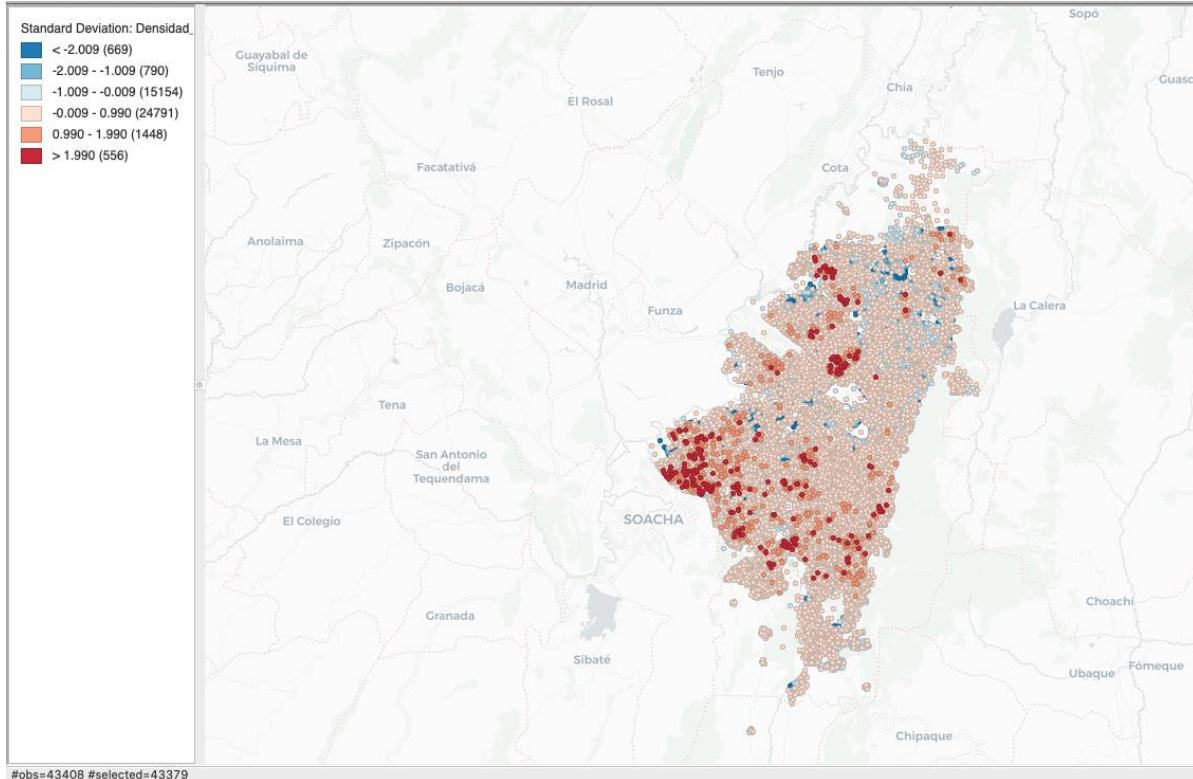
Segmentando estos valores, se encuentra que las localidades con mayor medición del indicador son parte de la localidad de Puente Aranda, Kennedy y Bosa. Se procede a realizar visualizaciones que expliquen estadísticamente la distribución espacial del indicador. La desviación estándar de los datos refiere de la idea de la medida de dispersión de los datos conforme a un valor esperado, por tanto las localidades con mayor desviación típica poblacional, se muestran a continuación:

Ilustración 50. Mapas comparativos de poblaciones con alta desviación estándar indicador DH.



Una gran parte de las manzanas de la localidad de "Bosa", presentan la concentración de manzanas con mayor desviación típica, es posible inferir que de acuerdo a los gráficos de cuartiles puede que para esta localidad se concentre una mayor magnitud del indicador pero existen una gran variación de registros de datos que presentan una gran diferencias en magnitudes, es decir es probable encontrar valor atípicos de inferiores magnitudes.

Análisis prospectivo de localidades con mediciones estadísticamente significativas del indicador :



Análisis de auto correlación espacial del indicador:

Realizando la modelación de la auto correlación de los valores de cada manzana conforme a la vecindad a partir de un gestor de matriz de pesos por contigüidad con el método KNN, obteniendo lo siguientes resultados:

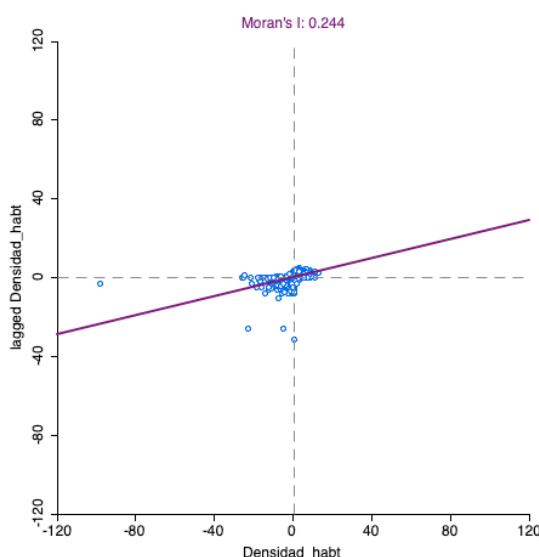
Tabla 31. Resumen de la creación de matriz de peso de contigüidad con el método KNN.

Property	Value
type	k-NN
inverse distance	true
power	1
symmetry	asymmetric
file	df_m_2do_ind_MGN20211125.gwt
id variable	field_1
distance metric	Euclidean
distance vars	coordinates
neighbors	4
# observations	43408
min neighbors	4
max neighbors	4
mean neighbors	4.00
median neighbors	4.00
% non-zero	0.01%

El máximo numero de vecinos que cumplen con el criterio de contigüidad y que por ende puede formarse una cadena de peso es de 4.

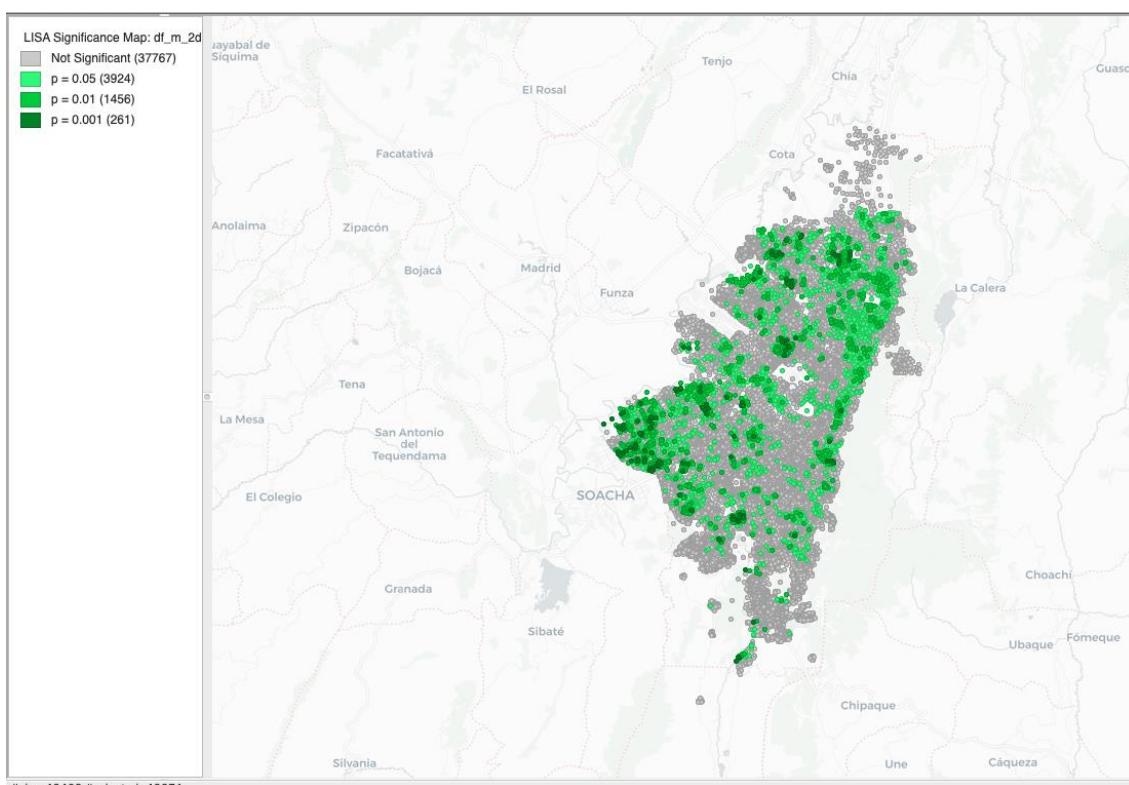
El primer estadístico a visualizar es el “Índice local de Morán” para el indicador :

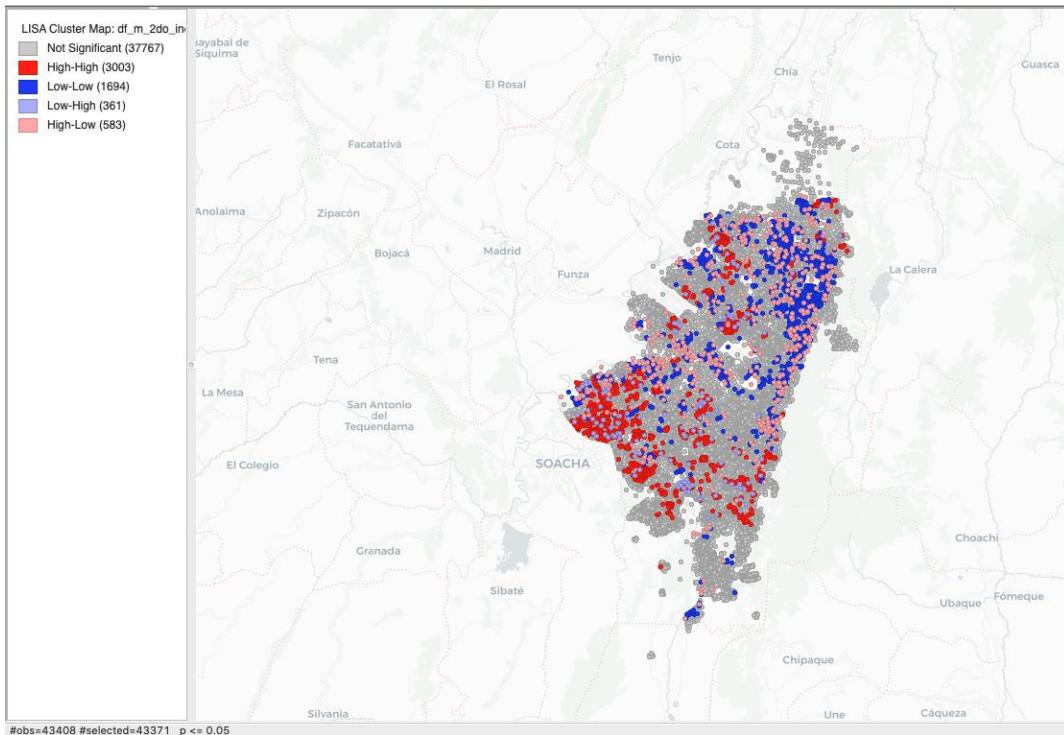
Gráfico 32. Gráfico del Índice de Morán Indicador DH.



Conforme a la medición del índice de morán, este se encuentra expresado bajo una magnitud de 0,244 aproximadamente 24,4%, indica que los datos en la totalidad presentan patrones completamente aleatorios en su distribución espacial debido a que se encuentra muy cercano a cero, además la mayor cantidad de datos se concentran en el tercer cuadrante y en el origen del plano, esto indica que la mayoría de manzanas poseen el valor esperado como medición a priori, y que la tendencia es que sean significativamente bajos.

Ilustración 51. Mapas de significancia y Clúster : Método de “Índice de Morán y LISA” 95% confiabilidad y 999 permutaciones.

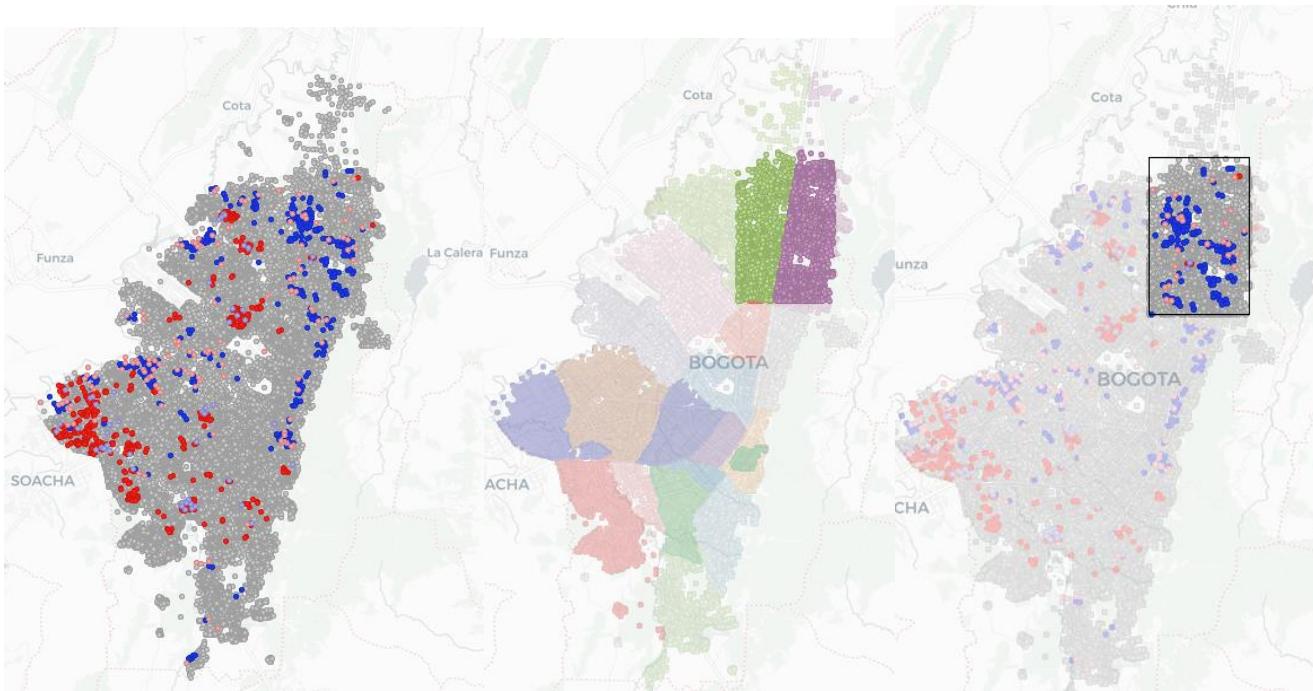




Se observa que para las diferentes entidades de la capa que están formando algún tipo de agrupación espacial de tipo clústeres de valores en alto superávit y bajo superávit, el nivel del valor para el indicador densidad habitacional de significación estadística corresponde al 13%. Se conforman los clústeres donde para un menor valor del “P-value” corresponde a la probabilidad de encontrar manzanas con valor correlacionadas con la entidad de contigüidad próxima o los vecinos próximos y vícera, o agrupaciones de manzanas que comparten características significativas y que ambas estén sujetas a condiciones significativas iguales de mediciones del indicador de valores altos se muestran con color “rojo” (Altos y altos). En el 95% de los casos se generan estas agrupaciones de zonas significativas.

Al aumentar el grado de confiabilidad del estudio de Hotspots, se obtiene entonces:

Ilustración 52. Mapas comparativos de significancia estadística con un 99% de confiabilidad ColdSpots



Para las manzanas en “Azul” (Método LISA) representadas anteriormente hacen referencia a que al generar todas las permutaciones posibles y luego de un proceso aleatorio, inminentemente estas zonas tendrían este patrón de agrupación con referencia a las condiciones de aleatoriedad de la distribución de los datos del indicador en la clasificación de significancia de los datos; estas se conocen como “ColdSpots”, representan observaciones con perfiles similares en términos estadísticos que están separados gradualmente y además que no representan un peso estadísticamente significativo que permita agruparse entre ellos. Las localidades “Usaquén” y “Suba”, presentan una alta concentración de manzanas que en sus agrupaciones reflejan valores de contigüidad bajos para la medición de las manzanas discriminadas en la anterior visualización y su vecindad refiere de valor bajos del indicador, es decir con un notable Déficit de hogares por lo que los clústeres en que los valores del indicador son bajos, figurando que el 99% de los casos las cantidades de viviendas son significativamente mayores a las cantidades de hogares.

Ilustración 53. Análisis de las manzanas con mediciones elevadas del indicador : DH.

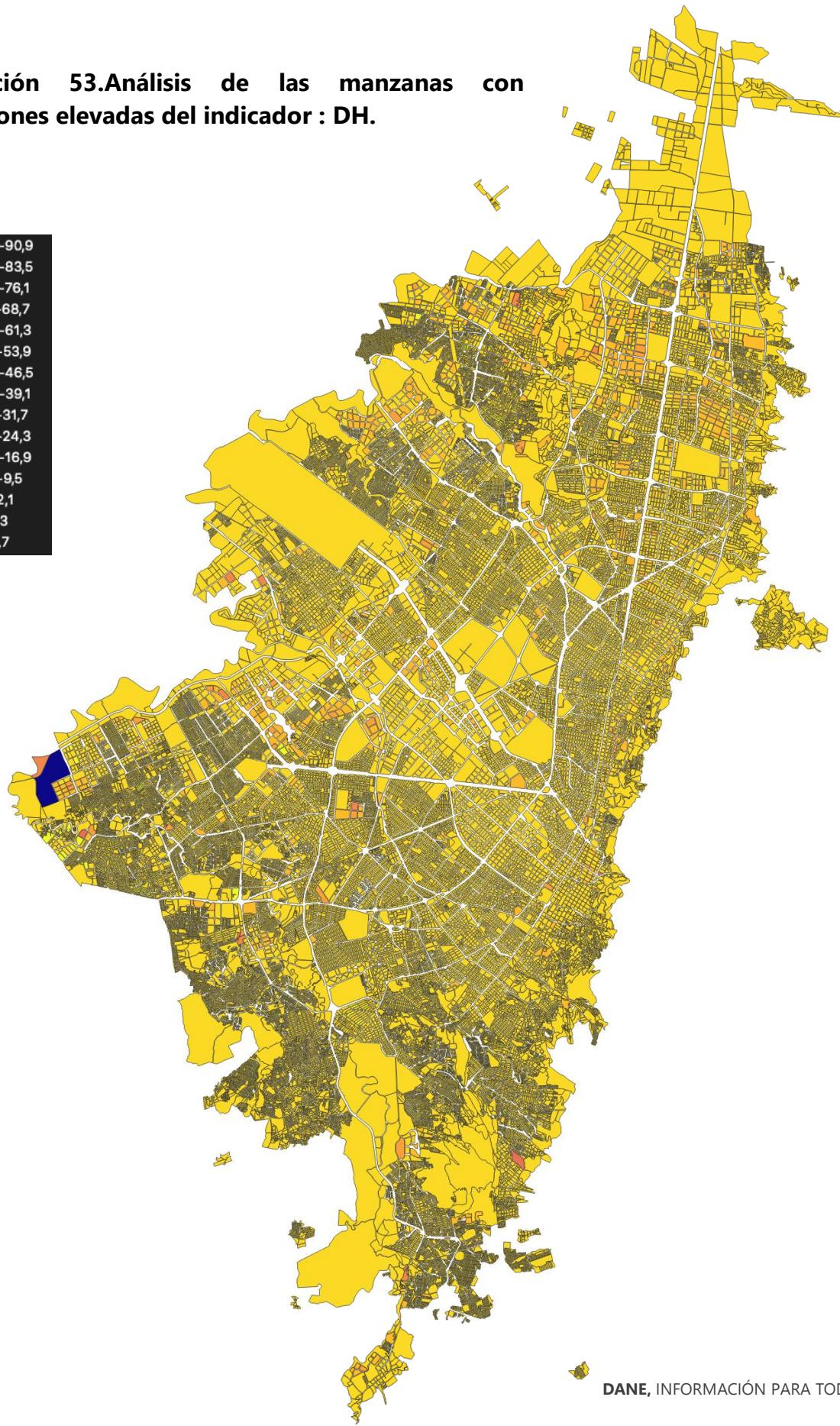
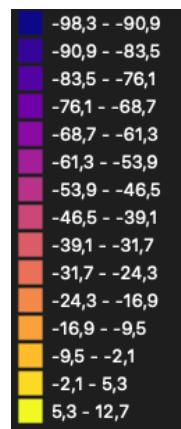
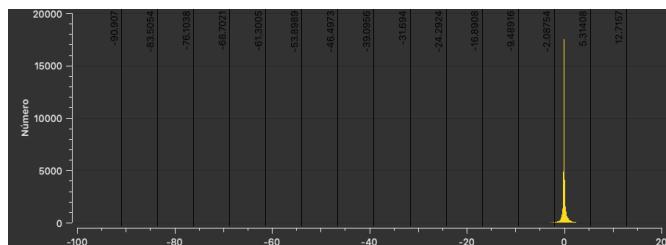


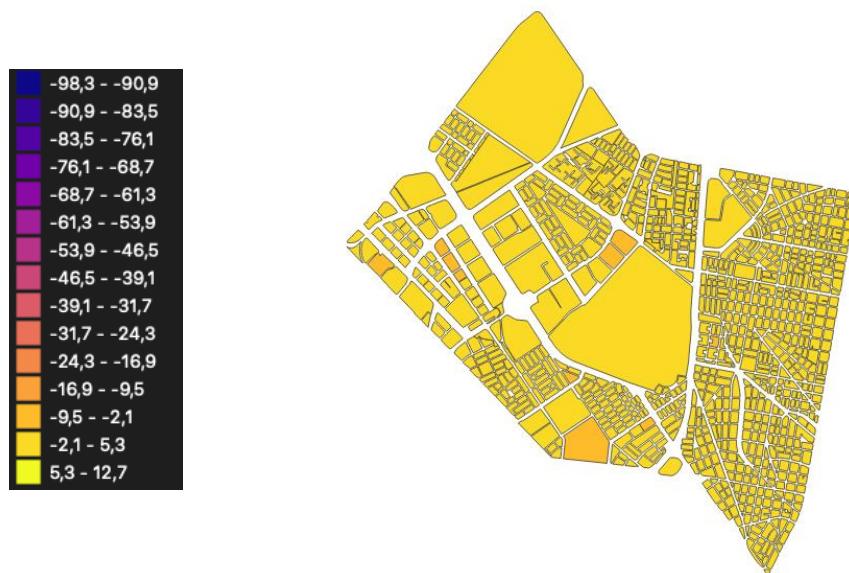
Gráfico 33. Gráfico de densidad del indicador DH.

En la clasificación del indicador por manzana en la ciudad de Bogotá DC, se presentan el **Gráfico 33**, una alta presencia de manzanas con colores para el valor 0, esto quiere decir que existen muchos registros de manzanas con altas de áreas que se encuentran en "Equilibrio", la cantidad de viviendas es igual a la cantidad de hogares.

Análisis de la localidad de Teusaquillo con mediciones del indicador :

Revisado de forma general las comparaciones entre los análisis Geoestadístico en la ciudad de Bogotá DC en cada una de las localidades, se procede a analizar las incidencias del valor de indicador en cada manzana de la localidad de "Teusaquillo" en la ciudad de Bogotá, procesando así clasificaciones para diferentes escalas del indicador medido:

- + En la localidad de Teusaquillo, se presentan manzanas con valores dentro de las escalas del [-5,5], cuya presencia de datos se encuentran en categorías tal como "Déficit", "Equilibrio" y "Superávit" distribuyéndose en toda la localidad.

Ilustración 54. Mapa localidad Teusaquillo con distribución geoespacial del indicador DCSP.

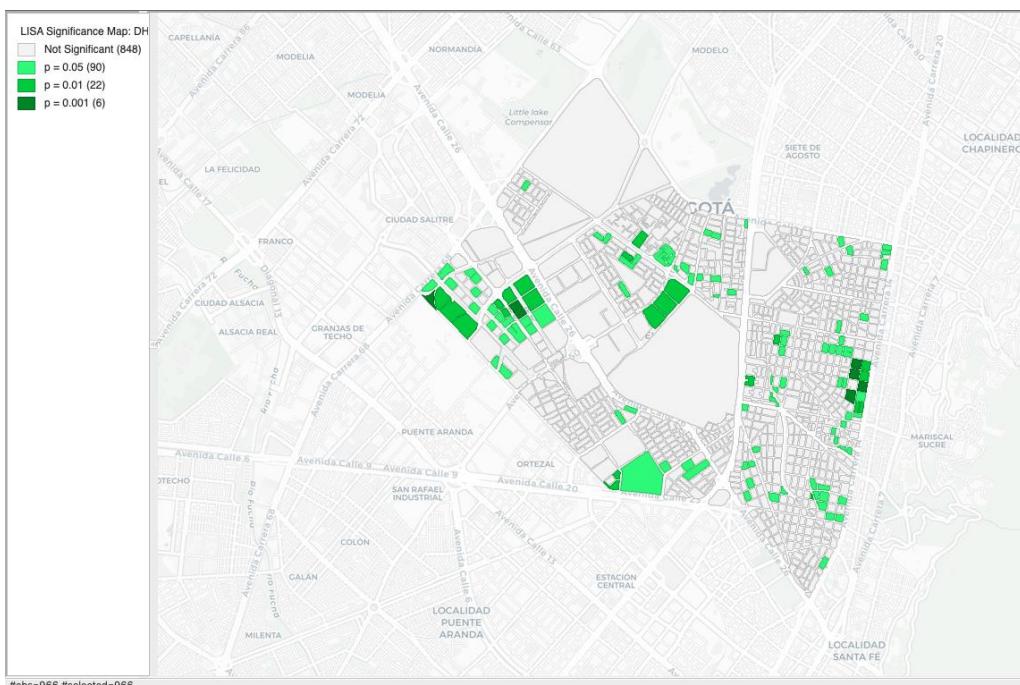
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE.

Tal como se muestra en la **Ilustración 54**, una selección en término medio de la localidad refleja la distribución de las manzanas conforme a sus colores dados por la clasificación (expresada en la leyenda), notándose así alta presencia de manzanas con colores "Amarillos" ubicadas en el rango de [-2.1,5.3], por tanto se concluye que la proporción de cantidad de viviendas con respecto a la cantidad de hogares es mayor y viceversa ya que el rango tiene contenido valores negativos, el cero y valores positivos.

Análisis prospectivo geoespacial de la localidad de Teusaquillo:

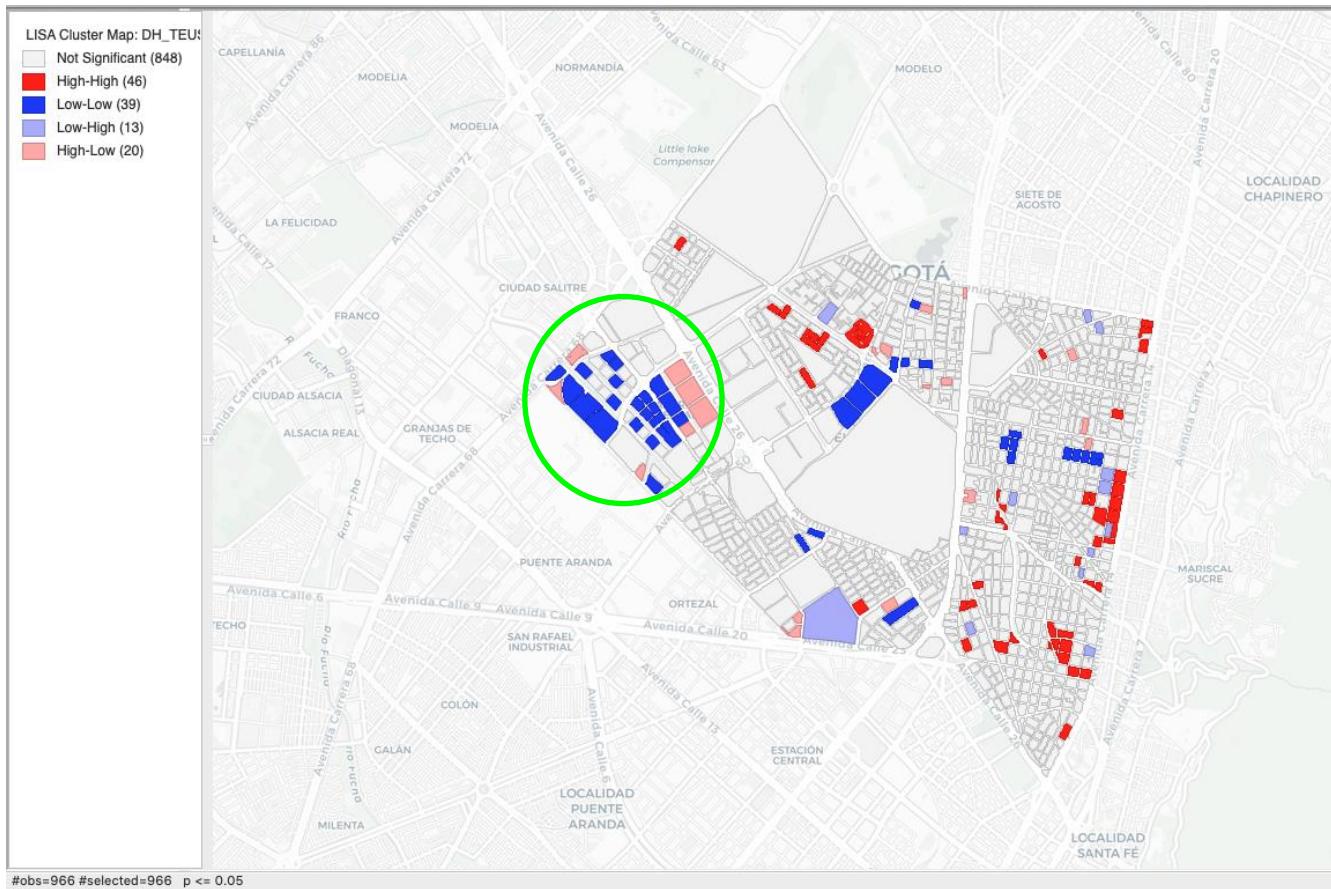
Las herramientas para análisis estadístico del indicador en cuestión, que permitan conocer las características de distribución de los datos expresando niveles de significancia estadística para cada manzana dentro de los perímetros de la localidad, se emplea la generación de mapas de significancia estadística a través del método del Índice de Morán. El proceso Geoestadístico del índice de morán mide la incidencia del valor univariante en función de un plano geoespacial, estableciendo clasificaciones de vectores espaciales que busca descubrir aquellas zonas en las que la auto-correlación espacial del indicador resulta ser elevado no solo debido a su condición estática de los datos sino que a través de procesos estadísticos que reflejan resultados excepcionales para el estudio de zonas en las que un grupo de manzanas resultan ser elevadas y están juntas unas con otras (o agrupadas), como también es posible encontrar zonas en las que no existe significancia de clústeres espaciales debido a que los valores varían mucho y no es posible encontrar agrupaciones de manzanas con valores ya sean elevados similares o inferiores similares unas manzanas con otras, teniendo en cuenta la vecindad.

Ilustración 55. Mapa de significancia método LISA, 95% confiabilidad para el indicador DH.



Las manzanas que muestra la **Ilustración 55**, son aquellas que sus valores no se deben a instancias aleatorias, sino que existe una fuerte influencia de su ubicación espacial para que estas tengan valores ya sean elevados del indicador o bajos del indicador de densidad habitacional.

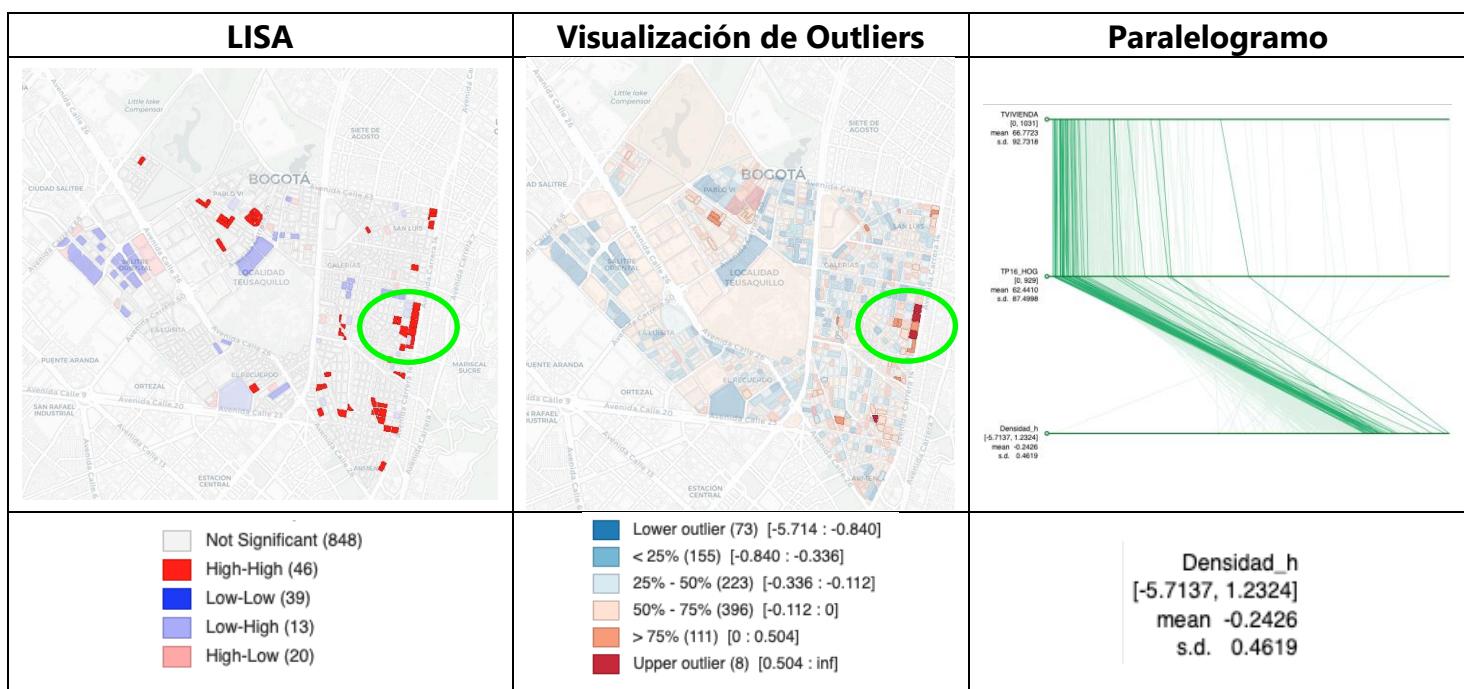
Ilustración 56. Mapa de clúster por el método LISA con peso de contigüidad KNN.



En el 95% de los casos es probable encontrar que las manzanas en los barrios de "Ciudad Salitre Oriental" (identificados en círculo verde) que se agrupen por tener mayor cantidad de viviendas que hogares, esto se debe a factores de la naturaleza de los predios en este caso las tres manzanas en el extremo izquierdo del barrio en cuestión corresponden a un conjunto residencial. Para las tres manzanas del lado inferior del barrio "La esmeralda" es de notarse que se presencian estos ColdSpots, es decir bajos niveles significativos en el data set sobre la densidad habitacional, existen menor densidad habitacional en estas tres manzanas que corresponden a conjuntos residenciales.

A continuación se analizan las manzanas con características del "HotSpot". El HotSpots hace referencia a la probabilidad de encontrar agrupaciones de manzanas con valores elevados del indicador, donde se encuentra mayor presencia de hogares que viviendas con un 95% de confiabilidad.

Tabla 32. Mapas de significancia : métodos LISA & Outliers en comparativa con Paralelogramo.



En el barrio de "La Esmeralda" se agrupan 7 manzanas con valores elevados del indicador, según el mapa de Visualización de Outliers se representa la incidencia de valores de desviación típica que para estas manzanas se encuentran en ">75%"(elevados), no obstante es de notarse que para el barrio las desviaciones típicas se encuentran entre el intervalo de [-0.336 , 0], lo que indica que las manzanas que representan valores significativamente altos de densidad habitacional no refieren de variabilidad en la población por tanto es plausible considerar que todas las manzanas seleccionadas e identificadas como "HotSpot" resultan ser altas por igual, a diferencias de las manzanas encontradas en el extremo sur-oeste de la localidad de Teusaquillo (identificado con el círculo verde), quienes a pesar de tener manzanas que forman clústeres de valores elevados es decir existe mayor población con hogares que con viviendas, la desviación típica es bastante alta encontrándose mucha variabilidad en esa zona, por tanto el indicador para todas las manzanas no se mantiene elevado o a los mismos niveles.

2.4 Construcción del indicador: “Garantía de desarrollo productivo”

El indicador de garantía de desarrollo muestra la configuración en los establecimientos y zonas de alta predominancia de actividad productiva. Busca medir con agrupaciones de variables, aquellos espacios en la ciudad que demuestran buscar provecho del suelo por el potencial y las zonas en las que se concentran espacios de incidencia económica que impulsa el desarrollo productivo. A continuación, se presentan las variables afines a su construcción:

Tabla 33. Variables afines al indicador Garantía de desarrollo productivo.

VARIABLE	INDICADOR	CLASE	DESCRIPCIÓN	Categoría original
TP9_2_1_MI	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades mixtas con uso	Conteo de unidades mixtas con uso no residencial industria	Industria
TP9_2_2_MI	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades mixtas con uso	Conteo de unidades mixtas con uso no residencial comercio	Comercio
TP9_2_3_MI	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades mixtas con uso	Conteo de unidades mixtas con uso no residencial servicios	Servicios
TP9_2_4_MI	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades mixtas con uso	Conteo de unidades mixtas con uso no residencial agropecuario, agroindustrial, forestal	Agropecuario, Agroindustrial, Forestal
TP9_2_9_MI	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades mixtas con uso	Conteo de unidades mixtas con uso no residencial sin información	Sin información
TP9_3_1_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Industria	Industria
TP9_3_2_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Comercio	Comercio
TP9_3_3_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Servicios	Servicios
TP9_3_4_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Agropecuario, Agroindustrial, Forestal	Agropecuario, Agroindustrial, Forestal
TP9_3_5_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Institucional	Institucional
TP9_3_6_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Lote (Unidad sin construcción)	Lote (Unidad sin construcción)
TP9_3_7_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Parque/Zona Verde	Parque/Zona Verde
TP9_3_8_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Minero-Energético	Minero-Energético

TP9_3_9_NO	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Protección/Conservación ambiental	Protección/Conservación ambiental
TP9_3_10_N	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso En Construcción	En Construcción
TP9_3_99_N	GARANTIA DE DESARROLLO > ALTO O BAJO	Conteo de unidades no residenciales	Conteo de unidades no residenciales con uso Sin información	Sin información

Indicador de garantía de desarrollo productivo.

La garantía de desarrollo productivo se considera elemental a la hora de establecer los parámetros excepcionales del crecimiento de la economía en el país, por lo que las variables asociadas a este indicador son consideradas optimas debido a las categorías en las que son clasificadas, permitiendo contrastar estas unidades espaciales con respecto a las zonas en las que las actividades estimulan la concentración e integración de tipo social.

Las unidades no residenciales hacen referencias a los espacios físicos independientes y privados para el uso particular y exclusivo de usuarios, cuya categoría enuncia que al ser no residencial refiere de la idea que los residuos generados en esta unidad son derivados de una actividad diferente de la residencial, por tanto las categorías mencionadas de tipo : "Industria", "Servicios", "Comercio", "Minero-Energético", etc , generan residuos y sus zonas son adecuadas para las operaciones productivas.

El indicador de garantía de desarrollo productivo mide las zonas en las que existen mayor concentración de unidades que participan en algún tipo de actividad productiva y las que no, considerando la idea que las actividades no productivas son aquellas en las que no existen transformación de insumos que permitan obtener bienes y servicios. Los niveles del indicador a un grado de 1 indican que en esa manzana existen totalidad de garantía de desarrollo, ya que se encuentra una presencia completa de manzanas cuyas unidades están siendo empleadas para el desarrollo económico de la ciudad, en cambio los niveles del indicador a un grado -1, indican que en esa manzana existen totalidad de garantía de recreación e improductividad.

Se agrupan las variables con mismas características y afines, es decir aquellas variables que comparten la misma categoría, sean estas productivas o no productivas. Adicionalmente se genera una variable representativa que integran las componentes de variables que no se consideran productivas con el nombre "TP9_NOPROD", como resultado de la sumatoria de ("TP9NOPROD = TP936NO + TP937NO + TP939NO") las variables que referencian según el diccionario ser "Lote baldío" ("TP9_3_6_NO"), parques o zonas verdes ("TP9_3_7_NO") y zonas de protección y conservación ambiental ("TP9_3_9_NO"). Simplificando significativamente las variables de 16 variables a solo 9 explicativas, se muestra el resumen estadístico de estas 9 variables, respectivamente:

Ilustración 57. Resumen estadístico de las variables agrupadas por categoría productiva.

TP9_INSTITUCIONAL	TP9_MINERO&ENER	TP9_CONSTRUCCION	TP9_INDUSTRIA	TP9_COMERCIO	TP9_SERVICIO	TP9_AGRO
Min. : 0.000	Min. : 0.00000	Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.00000
1st Qu.: 0.000	1st Qu.: 0.00000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.00000
Median : 0.000	Median : 0.00000	Median : 0.000	Median : 0.1095	Median : 1.000	Median : 0.000	Median : 0.00000
Mean : 0.1461	Mean : 0.002995	Mean : 0.1095	Mean : 0.5516	Mean : 2.807	Mean : 1.611	Mean : 0.01359
3rd Qu.: 0.000	3rd Qu.: 0.00000	3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: 3.000	3rd Qu.: 2.000	3rd Qu.: 0.00000
Max. :19.000	Max. :11.000000	Max. :21.0000	Max. :52.0000	Max. :180.000	Max. :173.000	Max. :39.00000
TP9_SIN_INFO	TP9_NOPROD					
Min. :0.00000	Min. : 0.000					
1st Qu.:0.00000	1st Qu.: 0.000					
Median :0.00000	Median : 0.000					
Mean :0.01193	Mean : 0.398					
3rd Qu.:0.00000	3rd Qu.: 0.000					
Max. :4.00000	Max. :111.000					

Las variables productivas son aquellas que hacen parte de algún sector de la economía, este caso las primeras siete variables seleccionadas con el recuadro rojo en la **Ilustración 57**, se busca entonces la sumatoria de todas estas variables para obtener el total de unidades que son consideradas productivas almacenadas en una sola variable llamada "PROD".

El indicador presenta una relación explicativa de los valores obtenidos para las poblaciones consideradas productivas y las poblaciones consideradas no productivas, para la cual la diferencia representa la predominancia del tipo de población en cada manzana, seguida de la fracción que muestra el peso de esa predominancia sobre el total de la población delimitada en estas 16 variables ya agrupadas.

Ecuación 20. Indicador Garantía de desarrollo 1 (productivo).

$$\text{Garantia desarrollo 1} = \frac{\text{PROD} - \text{NOPROD}}{\text{PROD} + \text{NOPROD}}$$

Los resultados de los resúmenes estadísticos de estas variables plantean que el máximo registro de unidades productivas en manzanas es de 217, un 95% mayor que el registro máximo de las unidades consideradas no productivas, y que para el calculo del indicador se obtuvieron 9623 manzanas que al aplicar la ecuación del indicador dieron como resultados NA's debido a las consideraciones del los valores del indicador (ya sea se haya dividido cero entre cero).

Ilustración 58. Resumes de variables empleadas para el indicador garantía de desarrollo 1.

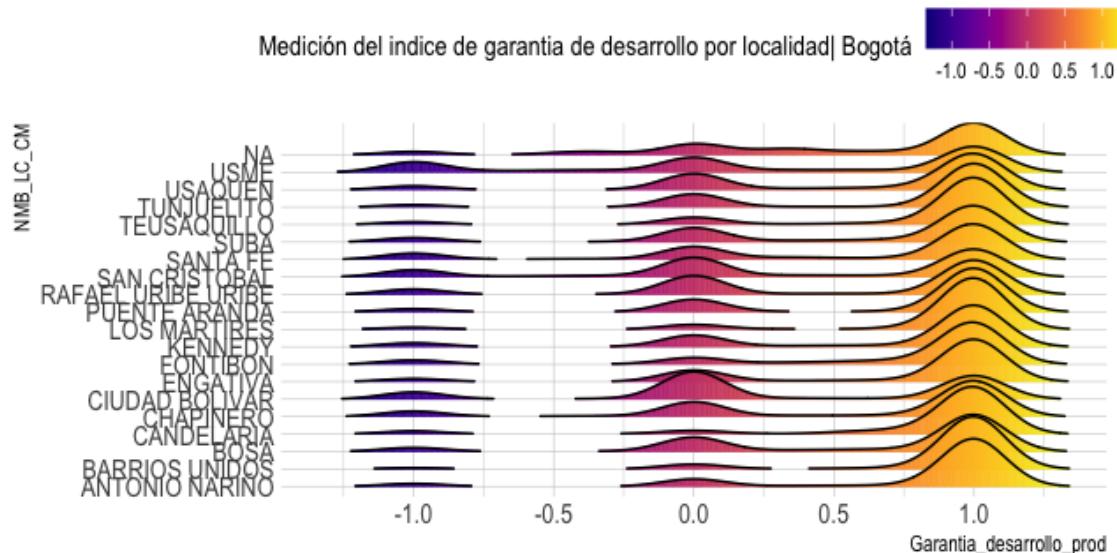
PROD	NOPROD	Garantia_desarrollo_1
Min. : 0.000	Min. : 0.000	Min. : -1.000
1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.951
Median : 2.000	Median : 0.000	Median : 1.000
Mean : 5.242	Mean : 0.398	Mean : 0.699
3rd Qu.: 7.000	3rd Qu.: 0.000	3rd Qu.: 1.000
Max. :217.000	Max. :111.000	Max. : 1.000
		NA's :9623

Análisis categórico exploratorio del indicador: "Garantía de desarrollo 1" (GD1).

El gráfico de densidad del indicador permite evidenciar que el volumen de datos que se concentran en el valor 1 del indicador es mayor que los que se concentran en el valor 0 (cero) y -1, entrando en detalle para la localidad de "Ciudad Bolívar", se presenta mayor cuantía de valores concentrados en 0 (cero) que en ambos extremos del indicador, esto indica que en esa manzana existen mayor balance entre las zonas en las que se consideran productivas que las que no se consideran productivas.

Gráfico 34. Gráfico de densidad del indicador GD1 por localidad

MGN- Nivel Manzana :Dataset 2021



Las categorías a considerar a la hora de clasificar la variable son las siguientes:

- [$-\infty, -1]$ → "Garantía nula"
- [$-1, -0.75]$ → "Garantía crítica"
- [$-0.75, -0.001]$ → "Baja Garantía"
- [$-0.001, 0.001]$ → "Balance"
- [$0.001, 0.75]$ → "Alta Garantía"
- [$0.75, 1]$ → "Completa Garantía"

A continuación se presenta el resumen estadístico de estas categorías:

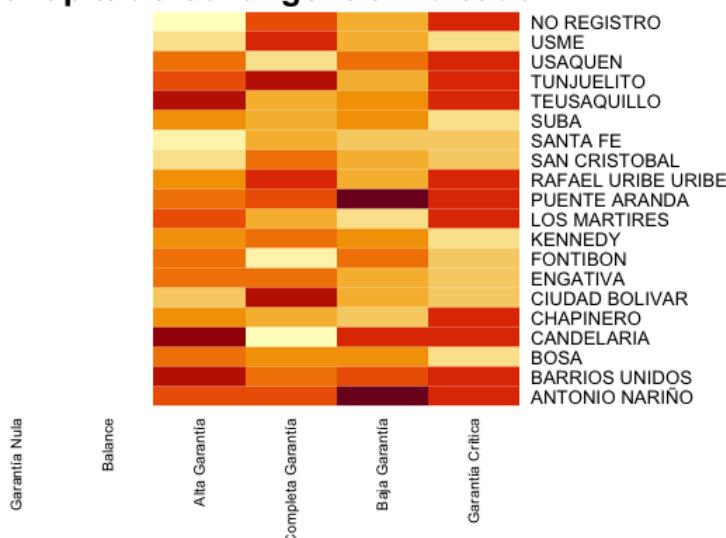
Tabla 34. Resumen de categorías indicador Garantía de desarrollo 1.

Garantía Nula	Garantía Crítica	Baja Garantía	Balance	Alta Garantía	Completa Garantía
3459	61	660	10577	2220	26431

La **Tabla 34**, indica qué el 60.88% de los datos se concentran en la categoría "Completa Garantía", más del 60% de las manzanas en general cuenta con una relación 1 a 1 de unidades productivas con respecto al total de unidades productivas e improductivas, es decir, que en esas manzanas con ausencia de unidades improductivas.

Ilustración 59. Mapa de calor del indicador Garantía desarrollo 1.

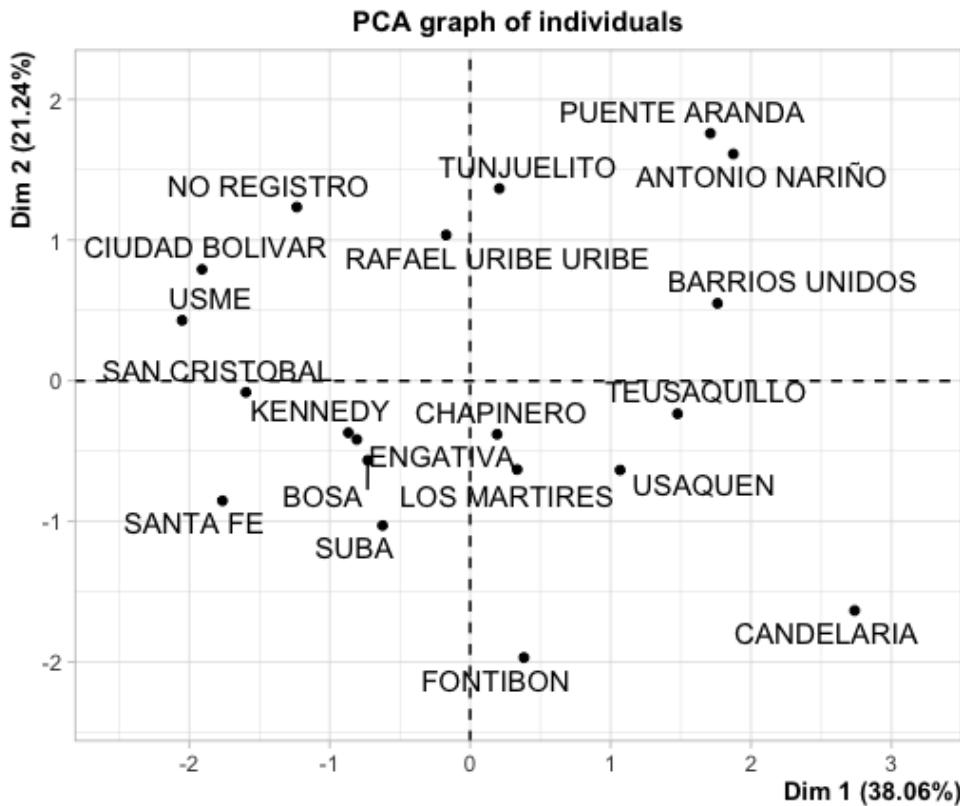
Heatmap tabla contingencia indicador



En el mapa de calor de la **Ilustración 32**, la categoría de "Alta Garantía" refleja las celdas con mayor oscuridad y corresponden de valores mayores o más elevados del indicador, así como se presenta para la localidad de "Teusaquillo", "Candelaria" y "Barrios Unidos", representando los valores en promedio más significativos de concentración en esta categoría. En las localidades de "Antonio Nariño" y "Puente Aranda" se aprecia que existen mayor cantidad de unidades no productivas debido a que existen valores en promedio para cada manzana que ascienden los valores en cada localidad en la categoría "Baja Garantía".

Efectuando un análisis de componentes principales, se obtiene que a través de dos componentes principales es posible explicar el 59,3% de la información de los datos.

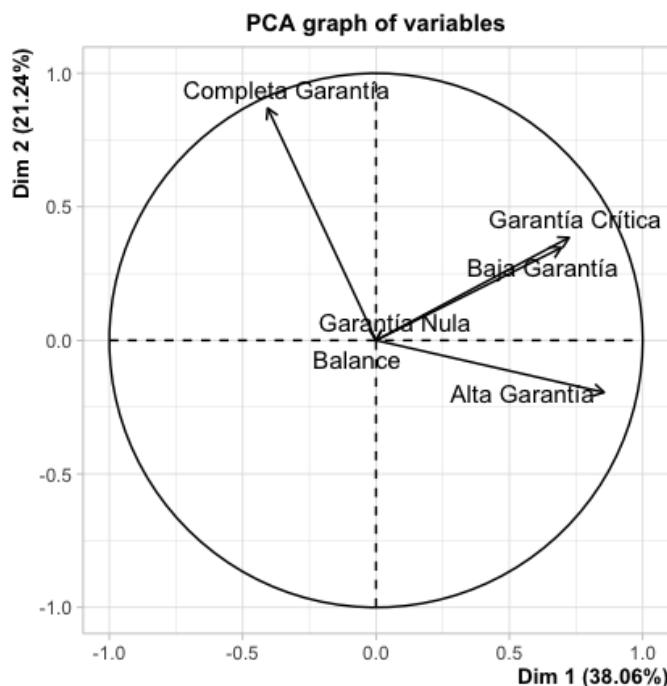
El método espacial muestra las ubicaciones por localidad dentro de las orientaciones en un plano bidimensional que estratifica con factores los perfiles de las localidades en cada categoría.

Diagrama 14. PCA Método Gráfico Espacial : GD1

Según el **Diagrama 14**, se muestran las localidades que representan el peso de participación de cada una en función del indicador de "Garantía de desarrollo productivo" ubicándose cada localidad en un extremo de cuadrantes construidos por las dos componentes, de tal forma que se contemple que componente contiene mayor número de localidades y que grado es explicativa cada localidad. Se evidencia la presencia de localidades que se encuentran muy dispersas en el método gráfico del PCA, donde la localidades de "Teusaquillo" y "Usaquet" se encuentran muy cerca y en el mismo cuadrante, esto indica una alta correlación entre estas localidades y su cercanía refleja que tienen perfiles muy similares en dirección a la categoría de "Alta garantía de desarrollo" representado en el **Diagrama 15**. Por su parte de acuerdo al método grafico, las localidades en el 3er cuadrante no presentan tanta dispersión sino que existen mayor presencia de localidades que se agrupan debido que tienen perfiles particulares y no se direccionan a un solo tipo de categoría sino que las condiciones de la garantía de desarrollo se distribuye para todas las categorías, y que sus valores en promedio del índice de garantía de desarrollo están en dirección contraria a las categoría "Baja Garantía" y "Garantía

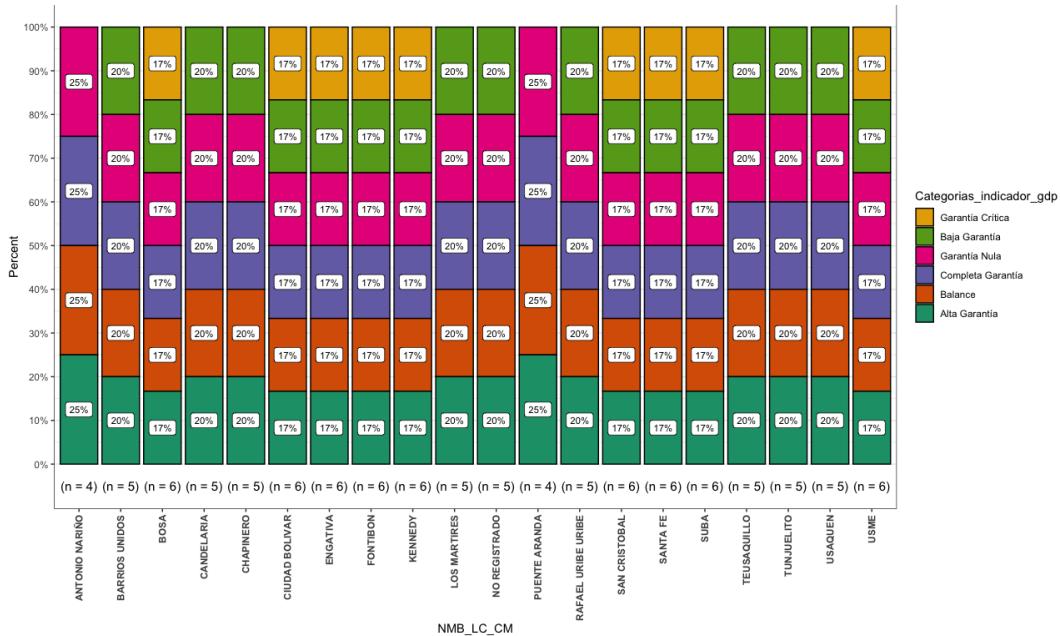
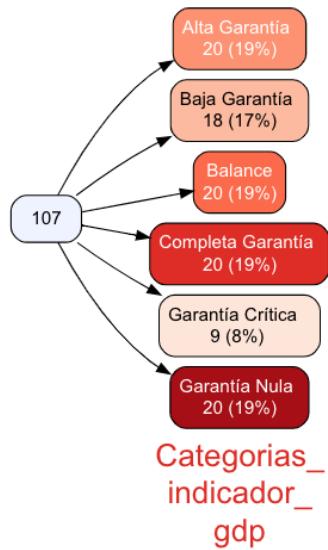
crítica”, si bien no representan alto peso en tener unidades productivas no se encuentran pesos ponderados de unidades no productivas en promedio por manzana.

Diagrama 15. PCA método circular.



En el método circular (**Diagrama 15**), la categoría “Completa garantía” va en diferente dirección que las otras categorías, si bien existen factores que influyen en que las categorías presenten ciertas similitudes de acuerdo a la concentración de datos, es fiable afirmar que las localidades que presenten mediciones significativas en “Completa garantía” serán estadísticamente diferentes a las que difieren en el cuadrante II (Superior izquierdo), adicionalmente en las categorías “Garantía crítica” y “Baja Garantía”, tienen exactamente la misma dirección (en el mismo ángulo de inclinación), el mismo sentido y magnitud, por lo que se afirma con más del 59% de la información que cualesquiera de las localidades se encuentren en una de estas dos categorías tiene el mismo nivel de participación de manzanas que se encuentren ambas categorías. A

nivel de detalle la categoría “Completa garantía”, presenta significativas diferencias en la composición de los vectores de datos en las tablas cruzadas y debido a que esta se encuentra en un solo cuadrante no se correlaciona con ninguna otra categoría

Gráfico 35. Gráfico de tablas cruzadas del indicador Garantía de desarrollo**Diagrama 16. Diagrama de árbol de las tablas cruzadas en el indicador Garantía de desarrollo.**

En el **Gráfico 35** como resultado del calculo porcentual de los valores de manzanas por categoría, se obtuvo que 9 localidades cuentan con igualdad de participación para las 6 (seis) categorías, estas refieren que las localidades en sus porciones de manzanas reflejan un intervalo de confianza en función de la proporción de las categorías mencionadas, por tanto es normal que dentro de estas localidades (Kennedy, Fontibón, Engativá, etc) 1 de cada 6 manzanas se

encuentren en cualquiera de las categorías. Existen 9 Localidades (Usaquén, Tunjuelito, Teusaquillo) que excluyen la categoría de "Garantía crítica", estas tienen participación del 20% en cada categoría a excepción de la crítica, esto asegurando un rango de proporciones de unidades por encima del 50%, lo que indica que en esas manzanas existen mayor presencia de unidades productivas. Las localidades con mayor balance de categorías productivas son las de "Puente Aranda" y "Antonio Nariño" debido a que el 75% de las manzanas se encuentran en las mejores categorías de garantía de desarrollo.

En el **Diagrama 15** se obtuvo para el calculo de 107 manzanas representativas que agrupan los promedios para cada categoría, diferentes participaciones, unos resultados de participación para cada categoría, respectivamente: La categorías "Alta garantía", "Balance", "Completa garantía" y "Garantía Nula" con 19% para cada una, "Baja Garantía" con un 17%, acumulando un total del 92% de las manzanas representativas, el restante se agrupa para las manzanas representativas con un 8% para la categoría "Garantía crítica".

Análisis prospectivo de localidades de la ciudad de Bogotá DC con mediciones estadísticamente significativas del indicador : "Garantía de desarrollo productivo":

Análisis de auto correlación espacial del indicador en la localidad de Teusaquillo.

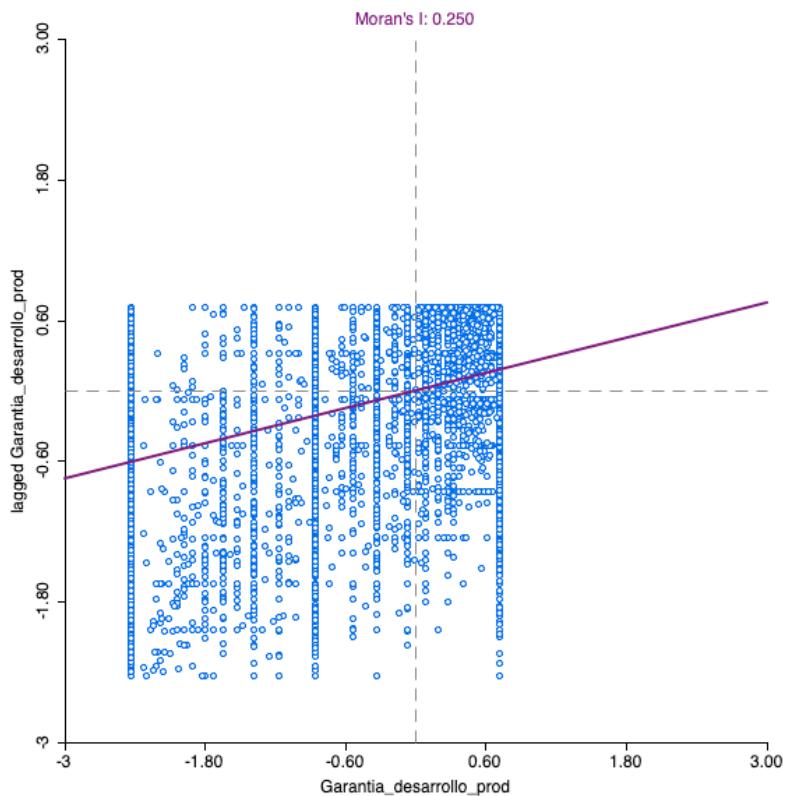
Realizando la modelación de la auto correlación de los valores de cada manzana conforme a la vecindad a partir de un gestor de matriz de pesos por contigüidad con el método KNN, se procede a un análisis prospectivo de la localidad de Teusaquillo, Bogotá DC.

Tabla 35. Resumen de la creación de matriz de peso de contigüidad con el método KNN.

Property	Value
type	k-NN
inverse distance	true
power	1
symmetry	asymmetric
file	df_m_3ero_1_ind_MGN20211202.gwt
id variable	X
distance metric	Euclidean
distance vars	coordinates
neighbors	4
# observations	43408
min neighbors	4
max neighbors	4
mean neighbors	4.00
median neighbors	4.00
% non-zero	0.01%

El máximo numero de vecinos que cumplen con el criterio de contigüidad y que por ende puede formarse una cadena de peso es de 4.

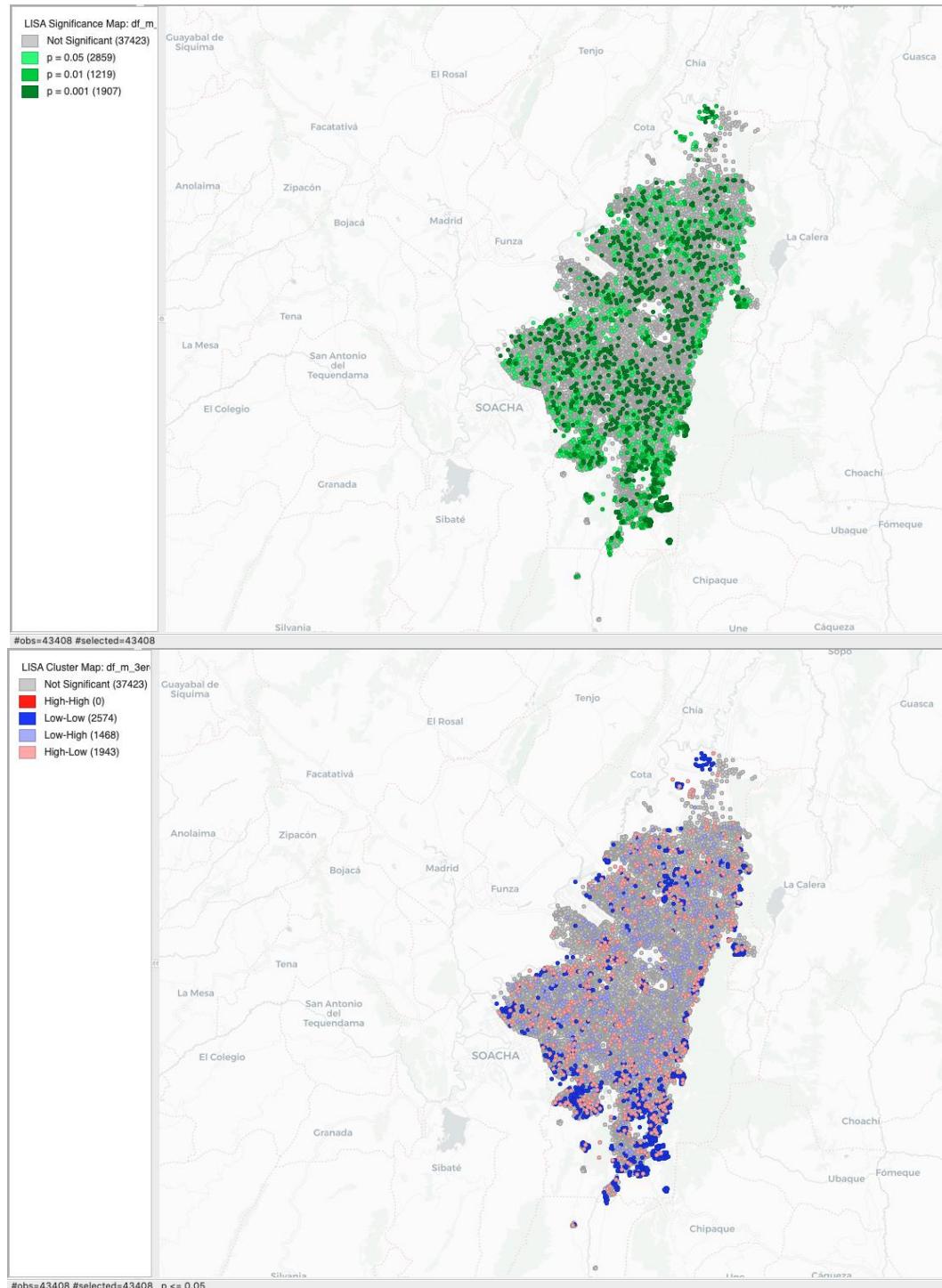
El "Índice local de Morán" para el indicador se explica a continuación:

Gráfico 36. Gráfico del Índice de Morán Indicador DH.

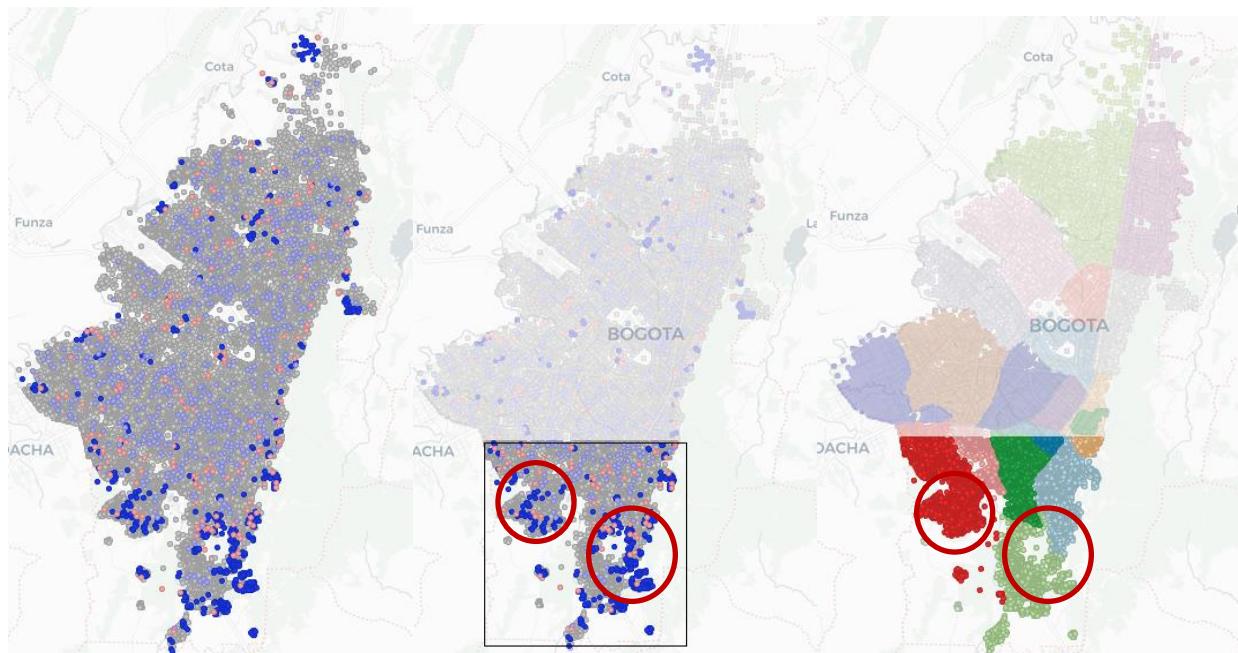
Conforme a la medición del índice de morán univariante del indicador, se encuentra expresado bajo una magnitud de 0,25 aproximadamente 25%, indicando que los datos en la totalidad presentan patrones completamente aleatorios y es posible aceptar la hipótesis que la distribución espacial de los datos es completamente aleatoria debido a que se encuentra muy cercano a cero, además se evidencia un patrón de concentración de datos organizados en un mismo grado de las escala en el IV cuadrante (inferior izquierdo), indicando la existencia de gran incidencia y dispersión de las manzanas improductivas rodeadas de manzanas que son improductivas.

Para las cuatro áreas del gráfico, corresponden a los 4 (cuatro) tipos de asociación espacial local posibles, entre una manzana y sus vecinos. En el cuadrante 1 (superior derecho) del gráfico del índice de morán, indica que existe una selección de regiones con alto valor del índice de garantía de desarrollo productivo, es decir existen manzanas con mayor cantidad de unidades productivas rodeadas de manzanas con las mismas características, en el cuadrante 2 (superior izquierdo) indican la presencia de manzanas con alta garantía desarrollo rodeadas de manzanas con baja garantía de desarrollo, el cuadrante 3 (inferior izquierdo) las regiones que se encuentran manzanas con bajo garantía de desarrollo productivo rodeadas de manzanas con baja garantía.

Ilustración 60. Mapas de significancia y Clúster : Método de “Índice de Morán y LISA” 95% confiabilidad y 999 permutaciones.



Se observa que para las diferentes entidades de la capa representada en el **Ilustración 60**, que están formando algún tipo de agrupación espacial de tipo clúster para los valores del indicador en diferentes categorías, se identifican las manzanas que se ajustan a los valores con diferentes clasificaciones. Según la leyenda del mapa de clústeres, existe mayor presencia de manzanas consideradas con zonas de baja productividad (azules) debido que contiene mayor cantidad de manzanas comparadas con todas las categorías del mapa a excepción de las "No significativas", llegando a contribuir con un 5,93% del total, la categoría "low-high" con un 3,38% y "high-low" con 4,47% (Manzanas productivas rodeadas de improductiva y viceversa). La probabilidad de encontrar manzanas correlacionadas con la entidad de contigüidad próxima o los vecinos próximas y vicersa, o agrupaciones de manzanas que comparten características bajas en significancias y que ambas estén sujetas a condiciones significativas iguales de mediciones del indicador de valores bajo se muestran con diferentes colores, para el color "azul" se presentan las manzanas de tipo baja productivas rodeadas de manzanas con la misma categoría sujetas a un grado de significancia igual que el 0,05, existe un alta presencia de zonas de color gris en el mapa de significancia, esto quiere decir que no existe pesos significativos en como esta distribuido el indicador como argumento suficiente para no crear agrupaciones en zonas en las que se pueda evidenciar altos niveles de garantía de desarrollo y estas ocupan un 86,21%. Adicionalmente, se presentan cero manzanas para la clasificación en un peso "High-High"(zonas en color rojo), es decir esas zonas ubicadas con agrupaciones de manzanas con alta garantía de productividad al momento de permutar 999 los valores del peso de contigüidad (vecindad) en el 95% de los casos.

Ilustración 61. Mapas comparativos de significancia estadística con un 99% de confiabilidad ColdSpots

Para las manzanas en "Azul" (Método LISA) representadas anteriormente en la **Ilustración 61**, conforme a las condiciones de aleatoriedad de la distribución de los datos del indicador cuando se muestran estos puntos "ColdSpots", estos hacen referencias a que las observaciones con perfiles similares en términos estadísticos están separados gradualmente y además que no representan un peso estadísticamente significativo que permita agruparse entre ellos, pero que en este caso para las localidades "Ciudad Bolívar", "San cristobal" y "Usme", se muestra una alta concentración de manzanas que en sus agrupaciones reflejan valores de contigüidad bajos para la medición de las manzanas discriminadas en la anterior visualización y su vecindad refiere de valor bajos del indicador, es decir con una notable baja garantía de desarrollo productivo se constituyen los clústeres en que los valores del indicador son bajos, es posible afirmar que en el 99% de los casos las cantidades unidades improductivas son significativamente mayores a las cantidades de unidades productivas en esas localidades mencionadas.

Ilustración 62. Análisis de las manzanas con mediciones elevadas del indicador : GD1.

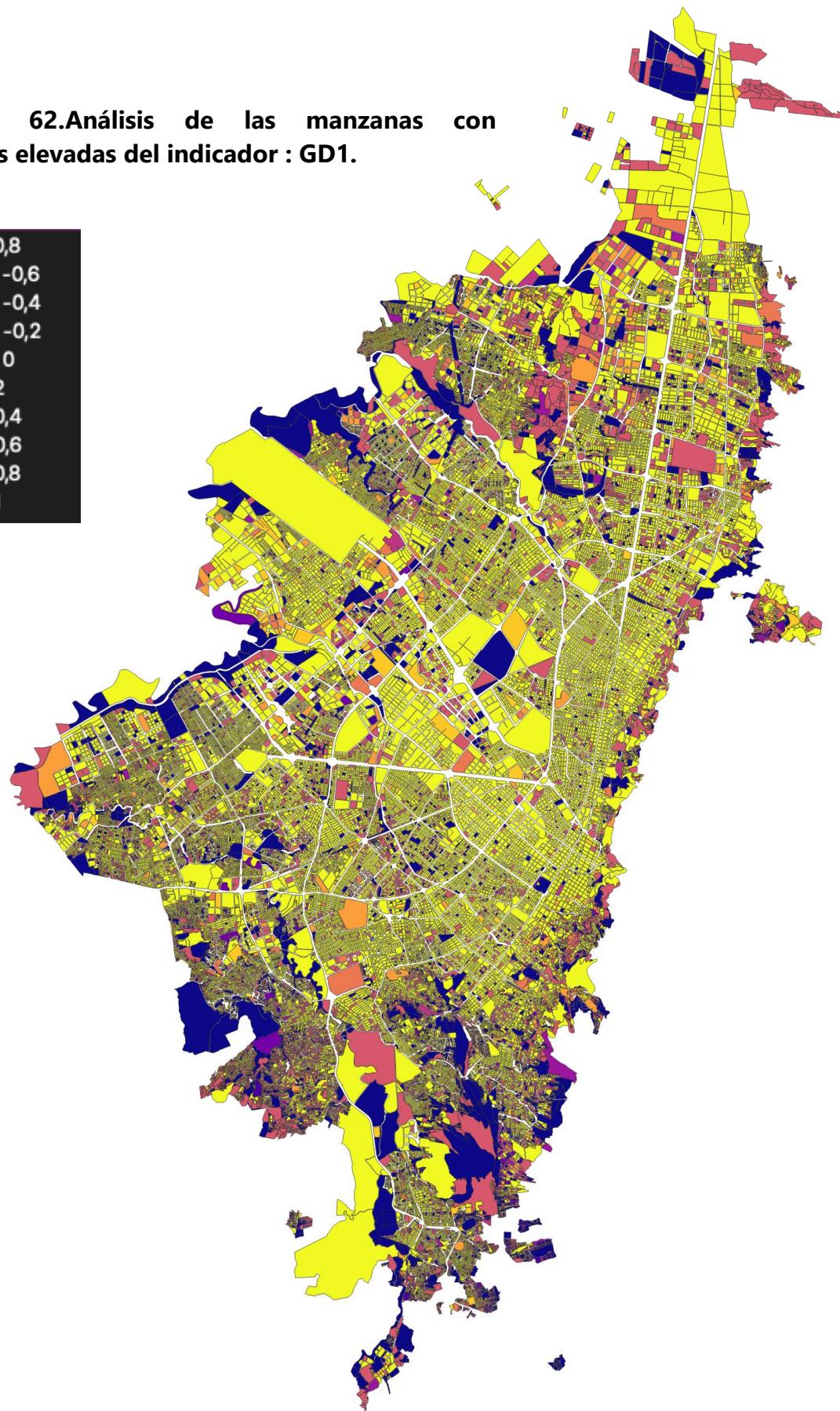
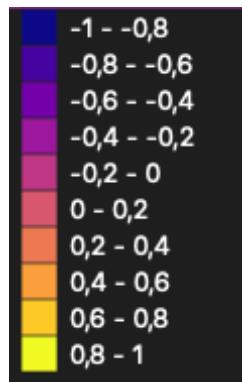
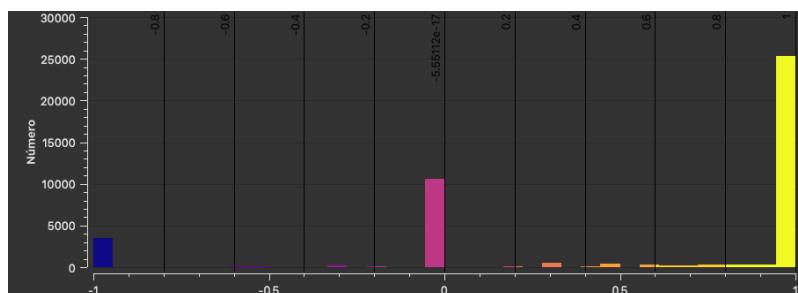


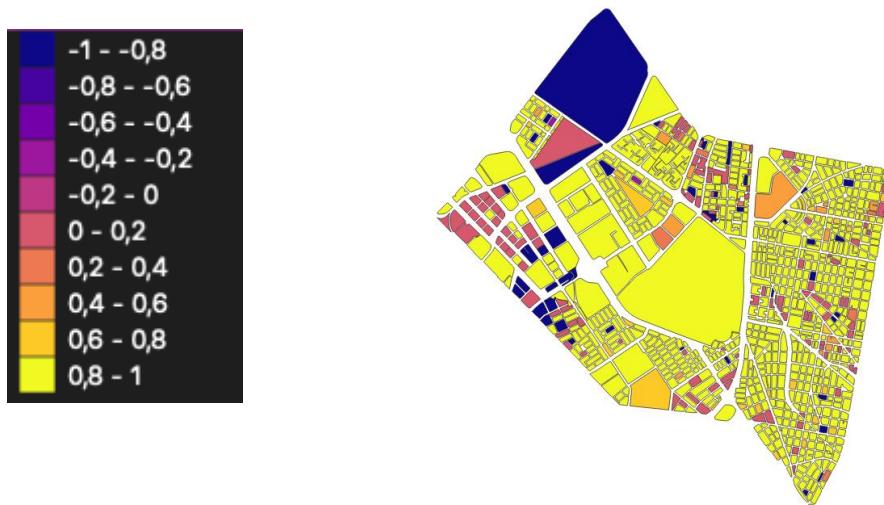
Gráfico 37. Gráfico de densidad del indicador GD1.

En el **Gráfico 37** se presenta la clasificación del indicador por manzana en la ciudad de Bogotá DC, es de notarse la distribución de las manzanas conforme a sus colores dados una alta presencia de manzanas con colores amarillo (valor 1), una media presencia de manzanas con color rosado (valor 0) y una inferior presencia con color azul (valor -1), esto quiere decir que existen muchos registros de manzanas con altas de áreas que se encuentra en "Completa garantía de desarrollo productivo", la cantidad de unidades productivas es igual al total de productivas más improductivas.

Análisis de la localidad de Teusaquillo con mediciones del indicador :

Un vez revisado de forma general las comparaciones entre los análisis Geoestadístico en la ciudad de Bogotá DC en cada una de las localidades, se procede a analizar las incidencias del valor de indicador en cada manzana de la localidad de "Teusaquillo" en la ciudad de Bogotá, procesando así clasificaciones para diferentes escalas del indicador medido:

Ilustración 63. Mapa localidad Teusaquillo con distribución geoespacial del indicador GD1.



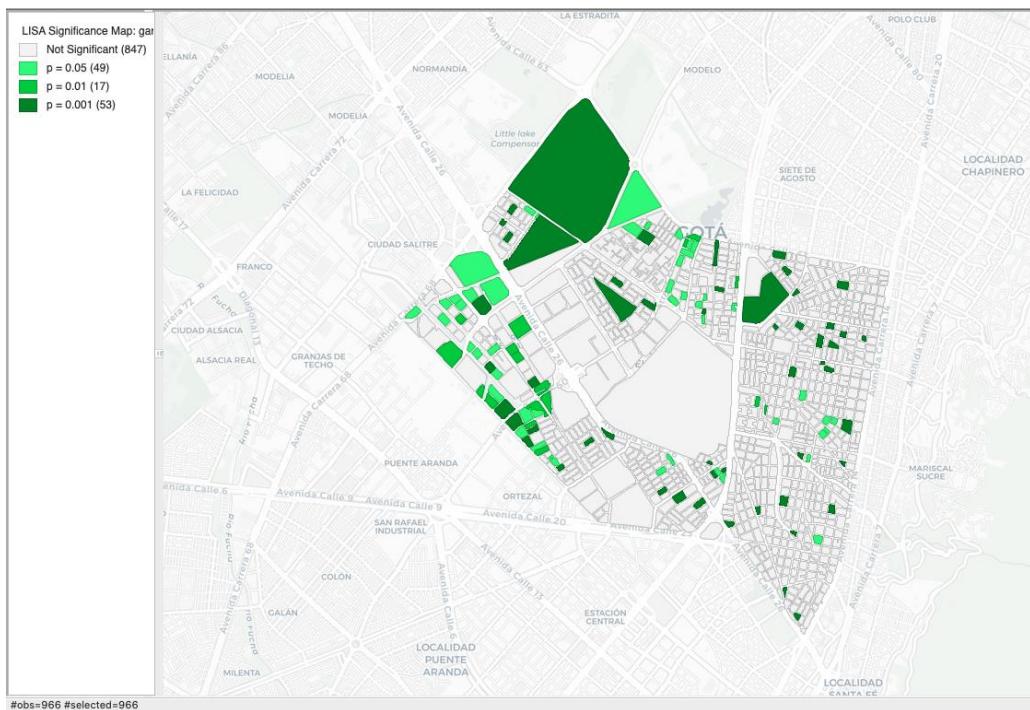
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE.

Para una selección en término medio de la localidad, se aprecia la distribución de las manzanas conforme a sus colores dados por la clasificación, notándose un alta presencia de manzanas con colores "Amarillos", estas ubicadas en el rango de [0.8,1], infiriendo una participación cumulativa de la proporción de unidades productivas con respecto a total de unidades discriminadas como el total de unidades en estudio.

Análisis prospectivo geoespacial de la localidad Teusaquillo:

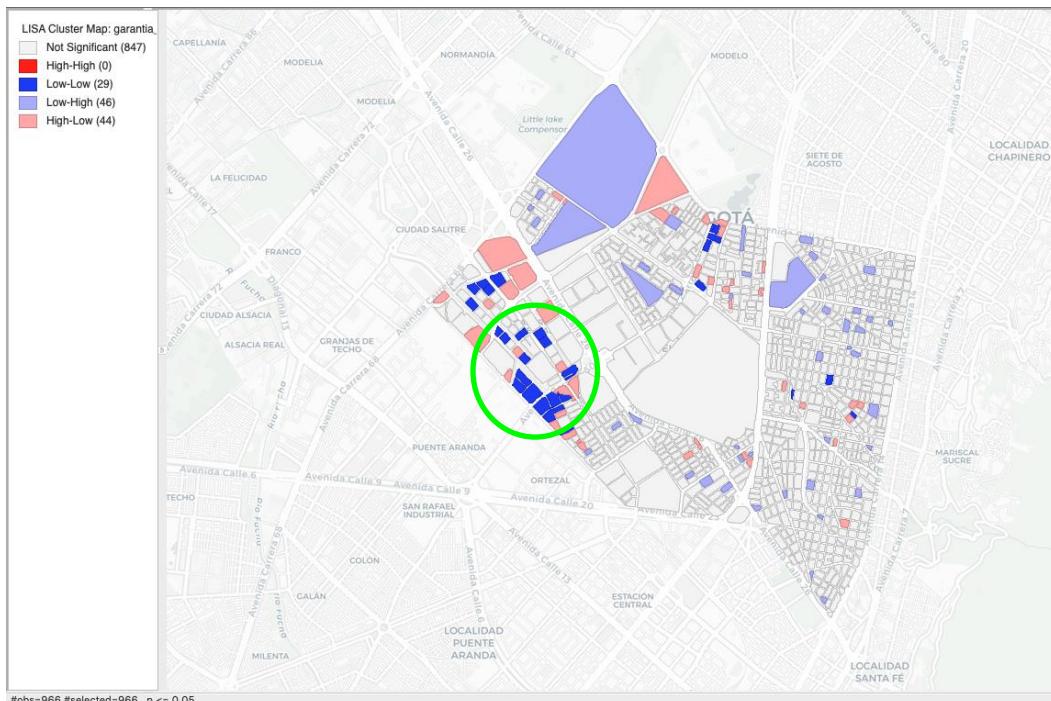
En la revisión geo-estadística del indicador en cuestión, para la localidad de Teusaquillo conforme a las características de distribución de los datos, es posible apreciar el desarrollo de los niveles de significancia estadística para cada manzana dentro de los perímetros de la localidad. Gracias a este método del Índice de Morán se generan los siguientes de mapas de significancia estadística.

Ilustración 64. Mapa de significancia método LISA, 95% confiabilidad para el indicador GD1.



Las manzanas que muestra el mapa, son aquellas que sus valores no se deben a instancias aleatoria, sino que existe una fuerte influencia de su ubicación espacial para que estas tengan valores ya sean elevados del indicador o bajos del indicador de garantía de desarrollo prod.

Ilustración 65. Mapa de clúster por el método LISA con peso de contigüidad KNN del indicador GD1.



En el 95% de los casos es probable encontrar que las manzanas en los barrios de "Ciudad Salitre Oriental" y "Quinta Paredes" (identificados en círculo verde 0) que se agrupen en algún tipo de clúster con valores con bajos niveles de garantía de desarrollo productivo, por tener mayor cantidad de unidades que no representan dentro del concepto de la garantía de desarrollo productivo alguna manifestación de los sectores de la economía, esto se debe a factores de la naturaleza de los predios. Este grupo de manzanas en color azul representan el "ColdSpot" de la localidad de Teusaquillo, no obstante para el barrio de la esmeralda es de notarse que solo existen 3 manzanas que se asocian a algún tipo de categoría, es de notarse que se presencian estas manzanas es posibles encontrar cantidades de unidades productivas rodeadas de unidades no productivas en niveles significativos de acuerdo a la lógica del indicador, no es posible descartar la posibilidad que en estas zonas particularmente en la manzana de la mitad las cantidades de unidades productivas asciendan a una ocupación con igual pronunciación que las no productivas. En general en la localidad de Teusaquillo, existen una gran participación de las manzanas que no se consideran representativa debido a que no se consideran que con sus medidas del índice de garantía de desarrollo productivo puedan formarse agrupaciones en zonas que refieran de valores significativamente altos o bajos a excepción de los mencionado previamente.

2.5 Indicador de garantía de desarrollo productivo por sector económico:

La garantía de desarrollo productivo por sector económico permite conocer aquellas zonas en las que existen concentración de unidades productivas de acuerdo a algún tipo de sector económico, es decir que por sus características permiten determinar su influencia en el desarrollo económico del país. De tal modo, que en la clasificación de los sectores de la economía de acuerdo al tipo de actividad que se lleva a cabo en una unidad productiva, se explican a continuación la pertinencia y generalidades de cada sector:

1) Sector 1 (SECT_1) : "Sector primario referente a actividades del agro.

Este primer sector contempla esas unidades de tierras que están siendo utilizadas para actividades de producción básica en la que predominan la extracción, producción y consecución de recursos naturales, de acuerdo a la transformación de estos recursos en materia prima no elaboradas y que comúnmente se destinan a ser aquellos insumos para el sector industrial. En este sector se agrupan la ganadería, la silvicultura, la caza, la pesca; por tanto este sector propicia la fuente primaria de recursos y es la conexión al eje fundamental de la economía de acuerdo al consumo. En esta primera variable entonces se agruparían aquellas observaciones que haga referencia de algunas de las actividades productivas anteriormente mencionadas.

2) Sector 2(SECT_2):"Sector secundario referente a las actividades industriales.

El sector industrial comprende todas aquellas actividades económicas de un país que tienen como misión la extracción y transformación de insumo en procesos netamente industriales en productos o mercancías que se implementan para satisfacer las necesidades directas de un mercado, lidiando con cadenas de abastecimiento, distribución conforme, que a día de hoy a evolucionado debido a su fuerte influencia y relación con el sector productivo.

3) Sector 3(SECT_3):"Sector terciario referente a las actividades de servicios.

El sector de servicios incluye las actividades que no producen mercancía, sino que por lo contrario buscan abastecer al mercado de bienes y servicios, ubicado entre los más importantes a la hora de corroborar el funcionamiento de la economía en general por su extensión con la demanda y oferta. Este sector aglomera las actividades comerciales, industrias hoteleras y de restauración, transporte, servicios financieros, telecomunicaciones, la educación, servicios profesionales y administraciones públicas.

A continuación se agrupan las variables de acuerdo a su sector, obteniendo el total por cada sector.

Ilustración 66. Resumen estadístico de variables por tipo de actividad afines al indicador de "Garantía de desarrollo por sector económico".

TP9_INSTITUCIONAL	TP9_MINERO&ENER	TP9_CONSTRUCCION	TP9_INDUSTRIA	TP9_COMERCIO	TP9_SERVICIO	TP9_AGRO
Min. : 0.0000	Min. : 0.000000	Min. : 0.0000	Min. : 0.0000	Min. : 0.000	Min. : 0.000	Min. : 0.00000
1st Qu.: 0.0000	1st Qu.: 0.000000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.00000
Median : 0.0000	Median : 0.000000	Median : 0.0000	Median : 0.0000	Median : 1.000	Median : 0.000	Median : 0.00000
Mean : 0.1461	Mean : 0.002995	Mean : 0.1095	Mean : 0.5516	Mean : 2.807	Mean : 1.611	Mean : 0.01359
3rd Qu.: 0.0000	3rd Qu.: 0.000000	3rd Qu.: 0.0000	3rd Qu.: 0.0000	3rd Qu.: 3.000	3rd Qu.: 2.000	3rd Qu.: 0.00000
Max. : 19.0000	Max. :11.000000	Max. :21.0000	Max. :52.0000	Max. :180.000	Max. :173.000	Max. :39.00000
PROD	Garantia_desarrollo_1	TOTAL_UNID				
Min. : 0.000	Min. :-1.0000	Min. : 0.000				
1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 1.000				
Median : 2.000	Median : 1.0000	Median : 2.000				
Mean : 5.242	Mean : 0.5437	Mean : 5.652				
3rd Qu.: 7.000	3rd Qu.: 1.0000	3rd Qu.: 7.000				
Max. :217.000	Max. : 1.0000	Max. :217.000				

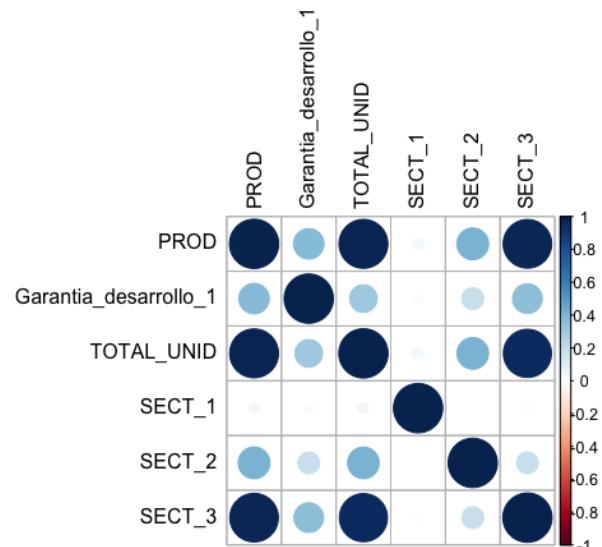
El primer sector de la economía corresponde solamente a la variable "TP9_AGRO"  , es decir aquellas unidades utilizadas para actividades de tipo agropecuaria, el segundo sector de la economía corresponde a las variables "TP9_MINERO&ENER", "TP9_CONSTRUCCION" y "TP9_INDUSTRIA"  , y el sector terciario con las variables de "TP9_INSTITUCIONAL", "TP9_COMERCIO" y "TP9_SERVICIO"  , tal como se observa en la **Ilustración 66**. Es de interés evidenciar el comportamiento de las medidas de estos valores referidos por sector económico en la ciudad de Bogotá, de tal modo que permita constatar hallazgos de instancias actuales de las localidades por ende se establece entonces un análisis categórico de las 3 variables y posteriormente un análisis factorial para determinar los pesos factoriales de cada uno para la consecución de una sola variable representativa que asocie los 3 sectores. A continuación se muestra el resumen estadístico para las tres nuevas variables.

Ilustración 67. Resumen estadístico integrando las tres variables de sectores económicos.

— Variable type: numeric —											
	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1 TP9_INSTITUCIONAL		0	1	0.146	0.595	0	0	0	0	19	
2 TP9_MINERO&ENER		0	1	0.00299	0.0850	0	0	0	0	11	
3 TP9_CONSTRUCCION		0	1	0.109	0.506	0	0	0	0	21	
4 TP9_INDUSTRIA		0	1	0.552	1.99	0	0	0	0	52	
5 TP9_COMERCIO		0	1	2.81	5.67	0	0	1	3	180	
6 TP9_SERVICIO		0	1	1.61	4.29	0	0	0	0	173	
7 TP9_AGRO		0	1	0.0136	0.278	0	0	0	0	39	
8 PROD		0	1	5.24	8.82	0	0	2	7	217	
9 Garantia_desarrollo_1		0	1	0.544	0.636	-1	0	1	1	1	
10 TOTAL_UNID		0	1	5.65	8.93	0	1	2	7	217	
11 SECT_1		0	1	0.0136	0.278	0	0	0	0	39	
12 SECT_2		0	1	0.664	2.08	0	0	0	1	52	
13 SECT_3		0	1	4.56	8.17	0	0	2	6	217	

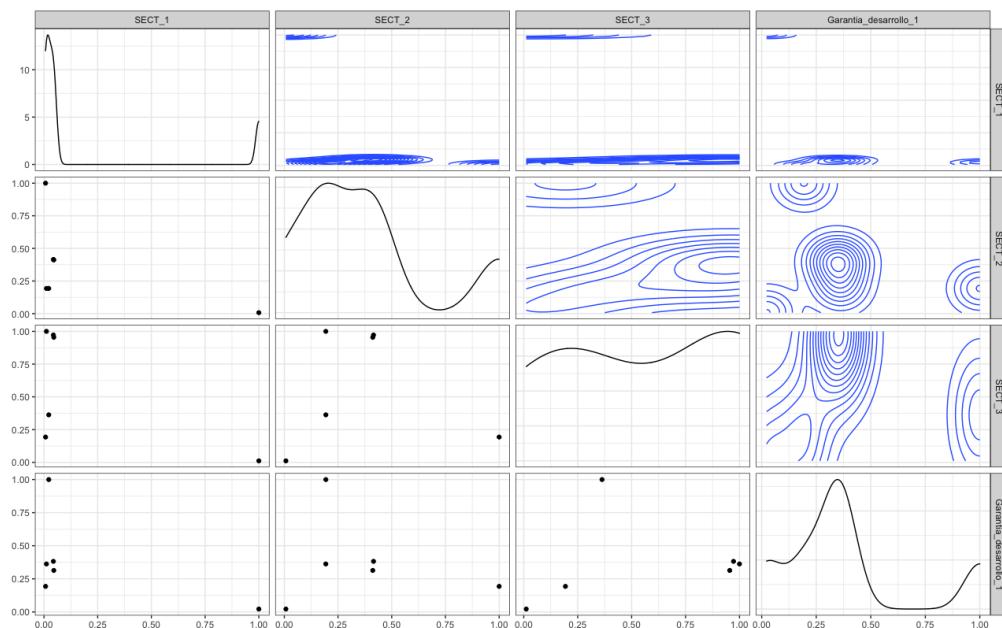
Las variables "SECT_1", "SECT_2" y "SECT_3", para cada sector económico es el resultado de la sumatoria las variables agrupadas por cada sector, y como es de apreciarse en el resumen estadístico anterior, la "sd" (desviación estándar) para el sector 3 (de servicios) es la que mayor variación tiene, es decir que esta variable refiere de la idea que cada manzana tendría una dispersión de magnitud diferente y que podrían encontrarse valores demasiados altos comparados entre otras manzanas comparados con las otras variables de sectores económicos, que reflejan una desviación estándar muy pequeña, lo que podría inferirse que no existe mucha dispersión en los datos y que las cantidades de unidades productivas en estos sectores por manzana, no varia mucho.

Ilustración 68. Matriz de correlaciones variables afines al indicador: "Garantía de desarrollo 2" (GD2).



Existe presencia de correlación entre las variables a excepción del sector 1, la cual demuestra un alto grado de independencia y nulidad por ser influenciada por las demás variables.

Ilustración 69. Correlograma de matriz de correlaciones de variables afines al indicador: GD2.



En el correlograma, es de evidenciarse el comportamiento del diagrama de densidad bidimensional que los intervalos en que los datos presentan correlación para las variables "SECT_2" y el indicador de "Garantía de desarrollo productivo" se concentran en los rangos de [0.25, 0.5] para el segundo sector de la economía y [0.25, 0.5] para el indicador GD1. Para el sector terciario las condiciones cambian dado que los intervalos de datos donde existe mayor concentración es de [0.5, 0.1] en función del indicador GD1 con un intervalo de valores de correlación de [0.25, 0.5], presentando este ultimo par mejor esquema de correlación.

Análisis de alfa de cronbach variables afines al GD2.

La fiabilidad de las variables dentro del procesamiento de los datos y las interpretaciones de acuerdo a las agrupaciones con el sector refieren de un valor métrico de confiabilidad de un 0.81, lo que quiere decir que las variables resultan de procesos fiables acordes a la realidad.

Ilustración 70. Resumen del alfa de cronbach variables afines al GD2.

```
Reliability analysis
Call: psych::alpha(x = dat1)
```

lower alpha	upper	95% confidence boundaries
0.81	0.81	0.82

Los resultados de la sensibilidad del alfa se presentan a continuación:

Tabla 36. Valores de sensibilidad del alfa de cronbach (raw_alpha)

	raw_alpha <dbl></dbl>	std.alpha <dbl></dbl>	G6(smc) <dbl></dbl>	average_r <dbl></dbl>	S/N <dbl></dbl>	alpha se <dbl></dbl>	var.r <dbl></dbl>	med.r <dbl></dbl>
PROD	0.6601183	0.6269676	0.7928115	0.2515790	1.680732	0.0006320058	0.08393618	0.1932028
Garantia_desarrollo_1	0.8419652	0.7720329	0.8486797	0.4038107	3.386598	0.0002725520	0.17508506	0.3023996
TOTAL_UNID	0.6668311	0.6376568	0.5288771	0.2603347	1.759814	0.0006114981	0.08830422	0.1932028
SECT_1	0.8482356	0.8431999	0.9488226	0.5181905	5.377545	0.0002837642	0.10338015	0.3969949
SECT_2	0.8313808	0.7759052	0.8620670	0.4091509	3.462397	0.0001261521	0.16995991	0.3381808
SECT_3	0.6808760	0.6623999	0.7539866	0.2818243	1.962085	0.0005416762	0.08893886	0.2535483

El valor de "raw_alpha" más bajo corresponde a la variable "PROD", esto quiere decir que al eliminar esta variable la fiabilidad de las variables pierde equilibrio de aspectos asociados a la realidad de proveniencia de los datos, por tanto es una variable critica, como también lo es la variable del "SECT_3", la cual al ser eliminadas la confiabilidad de los datos pasa de 0.81 a 0.68, esto ubicando a la fiabilidad en una escala de "Bueno" a "Malo".

Análisis categórico exploratorio de "SECT_1", "SECT_2" y "SECT_3".

Debido a que las variables más importantes son "PROD" y "SECT_3", la cual corresponden a total de unidades productivas por manzanas y el total de unidades productivas asociadas al sector terciario de la economía, a continuación se relacionan cada sector con la variable "PROD" de forma que se pueda conocer la proporción que ocupa cada sector en el total.

Las operaciones se almacenan en nuevas variables:

Ecuación 21. Proporciones de cada sector económico en términos del total de unidades productivas.

$$SECT_1.\text{fracc} = \frac{SECT_1}{PROD}$$

$$SECT_2.\text{fracc} = \frac{SECT_2}{PROD}$$

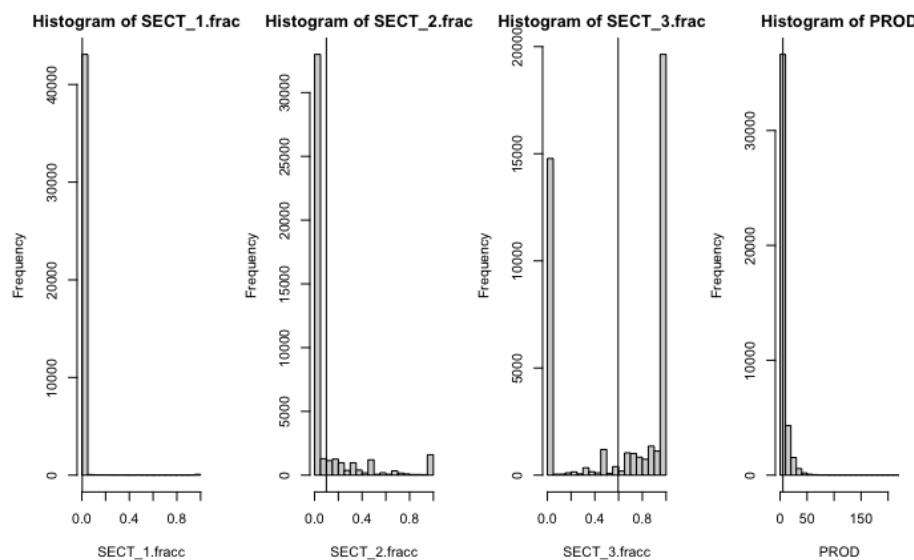
$$SECT_3.\text{fracc} = \frac{SECT_3}{PROD}$$

Ilustración 71. Resumen de las proporciones por cada sector económico.

SECT_1.fracc	SECT_2.fracc	SECT_3.fracc
Min. :0.00000	Min. :0.00000	Min. :0.0000
1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.0000
Median :0.00000	Median :0.00000	Median :0.8667
Mean :0.003348	Mean :0.09913	Mean :0.5962
3rd Qu.:0.00000	3rd Qu.:0.02395	3rd Qu.:1.0000
Max. :1.00000	Max. :1.00000	Max. :1.0000

Es de notarse en la **Ilustración 71**, que el sector terciario es que mayor tiene presencia y distribución de información en los cuartiles, esto demostrando que para este sector en la ciudad de Bogotá es el que mayor participación tienen debido que en promedio el 86,67% de las unidades productivas asociadas a este sector por manzana, y que el a partir del 70% de las manzanas registradas en este sector el 100% de las unidades corresponden a unidades destinadas a este sector, lo que es posible apreciar en el **Gráfico 38**.

Gráfico 38. Histograma de cada proporción de sector económico asociado y la variable PROD.



Las categorías a considerar a la hora de clasificar la variable son las siguientes:

- [− inf, 0.15] → "Deficiente participación sectorial"
- [0.15, 0.30] → "Equitativa participación sectorial"
- [0.30, 0.49] → "Equilibrada participación sectorial"
- [0.49, 0.51] → "Moderada participación sectorial"
- [0.51, 0.9] → "Alta participación sectorial"
- [0.9, 1] → "Perfecta participación sectorial"

A continuación se presenta el resumen estadístico de estas categorías:

Tabla 37. Resumen de categorías de las variables a fines al GD2.

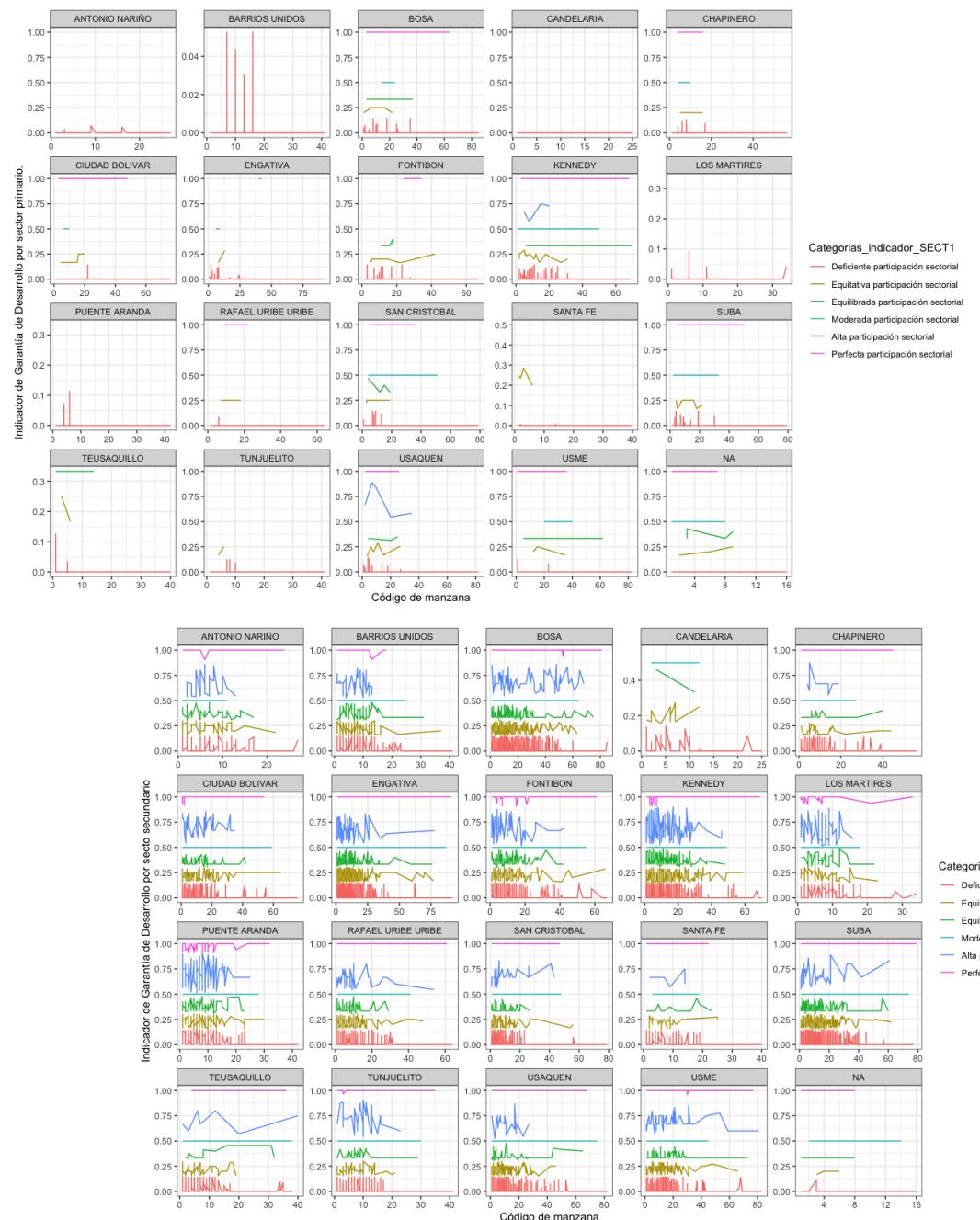
Categorias_indicador_SECT1	Categorias_indicador_SECT2
Deficiente participación sectorial :43172	Deficiente participación sectorial :35423
Equitativa participación sectorial : 71	Equitativa participación sectorial : 2600
Equilibrada participación sectorial: 36	Equilibrada participación sectorial: 1620
Moderada participación sectorial : 32	Moderada participación sectorial : 1123
Alta participación sectorial : 17	Alta participación sectorial : 1003
Perfecta participación sectorial : 80	Perfecta participación sectorial : 1639
Categorias_indicador_SECT3	
Deficiente participación sectorial :14866	
Equitativa participación sectorial : 316	
Equilibrada participación sectorial: 666	
Moderada participación sectorial : 1140	
Alta participación sectorial : 5658	
Perfecta participación sectorial :20762	

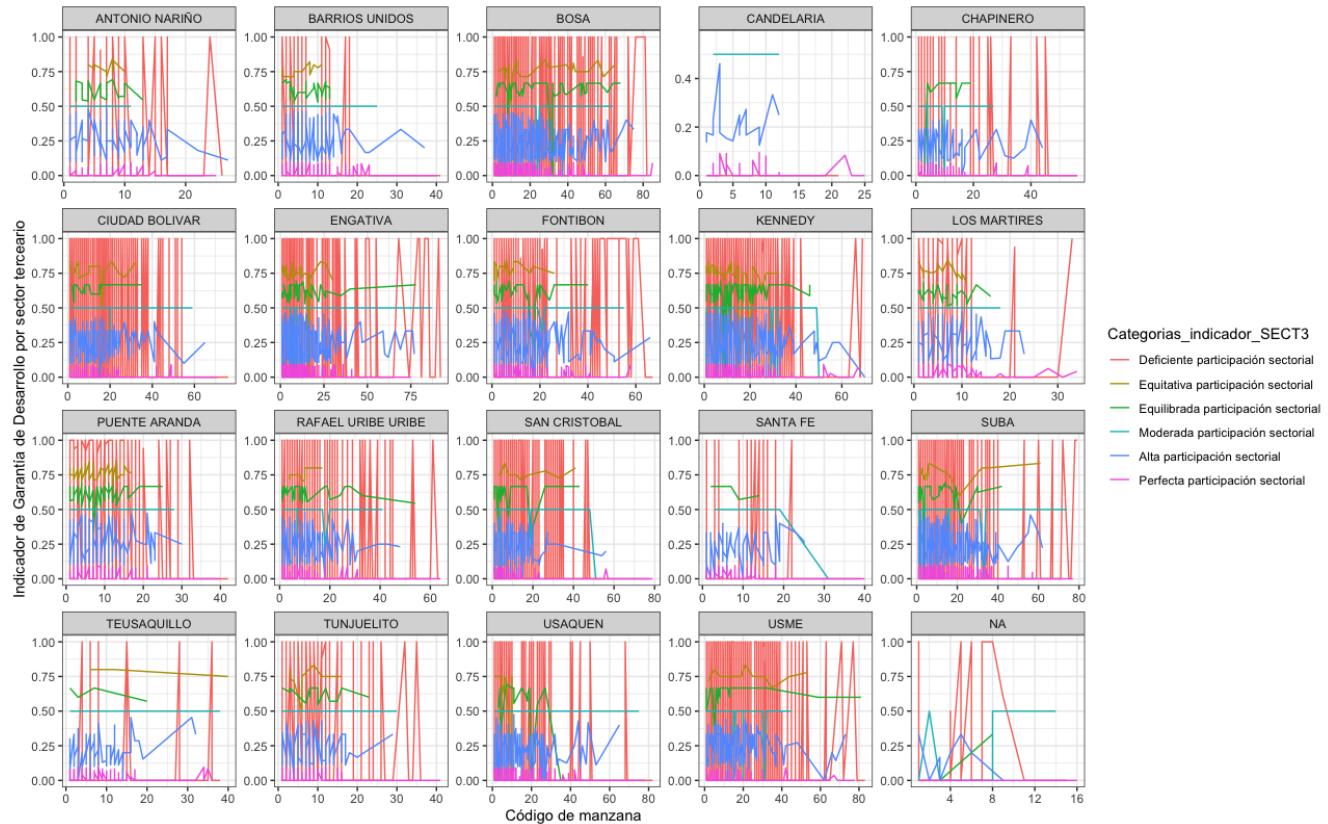
Para las categorías en el sector 1, solo 0.1843% de las manzanas se categorizan en perfecta participación sectorial, el 0.1635% de las manzanas se encuentra en un categoría de equitativa participación sectorial esto infiere en la idea que al menos existe un participación del 15% pero no mayor al 30% en cada manzana, en cambio para la categoría de deficiente participación sectorial es del 99.45%, la mayor cantidad de manzanas para este sector no confieren ninguna participación. En el sector 3, se presencia un 47.83% de participación de manzanas que se encuentran en perfecta participación sectorial o que más del 90% de las unidades productivas en esa manzana están destinadas a alguna actividad asociada a los servicios.

Análisis por manzana de cada sector económico a fin del GD2.

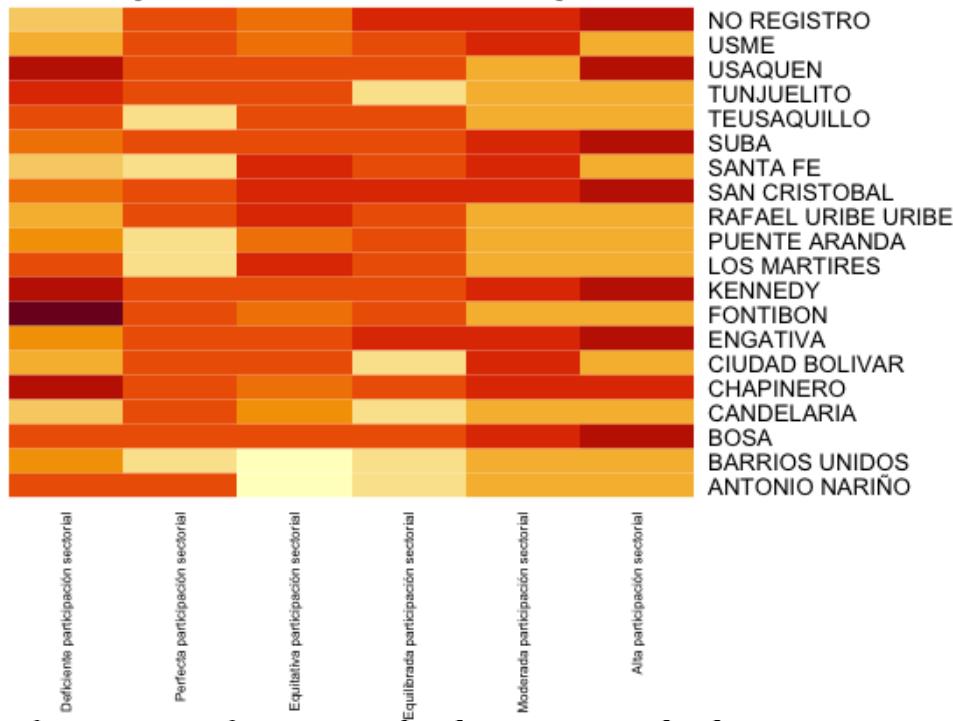
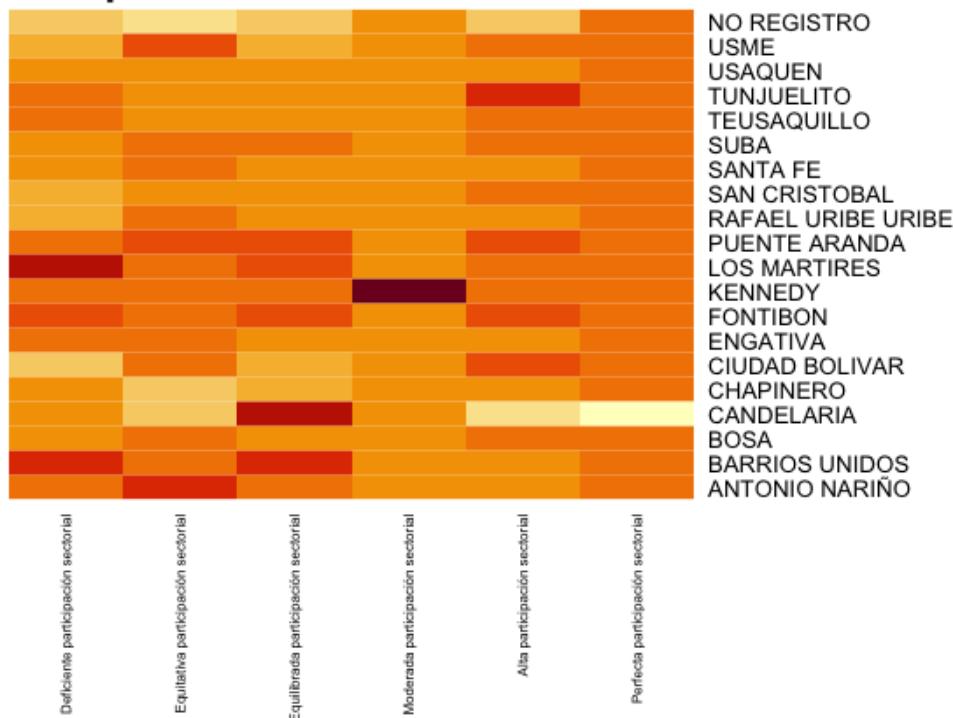
Las categorías para cada manzana permite conocer entonces en que localidades existen registros contundente de las categorías discriminadas y de acuerdo a las manzanas enumeradas en el data set como es están equipadas estas manzanas y su comparación en cada localidad.

Tabla 38. Comparación de categorías en función de los códigos de manzanas por localidad de los sectores económicos calculados.

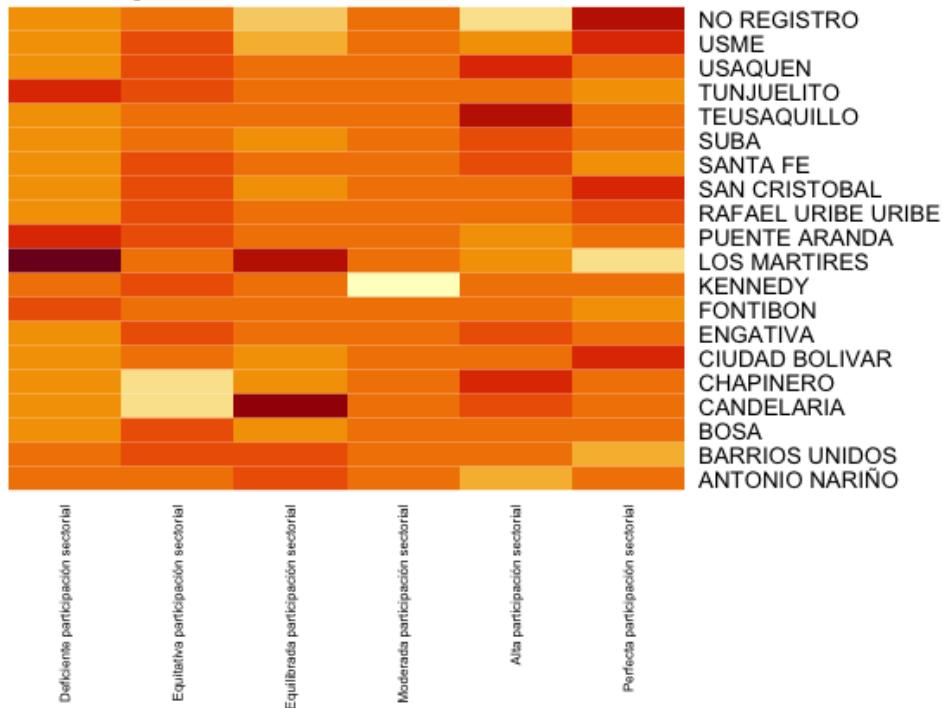




Las localidades "Bosa", "Kennedy", "San Cristóbal" y "Usaquén" para el primer sector, son las que más presencia de categorías devienen teniendo en cuenta que para cada localidad existe perfecta participación sectorial. De acuerdo al sector 3, las localidades con mejor condición del indicador de acuerdo a su distribución en diferentes categorías para cada código de manzana son: "Teusaquillo", "Santa Fe" y "Candelaria", para estas localidades la presencia de deficiencia de participación sectorial del sector secundario es baja. En la revisión del sector 2, la localidad "Candelaria" es la que menor cantidad de registros de manzanas con unidades productivas de tipo industrial tiene y las que tiene se ubican en las categorías de deficiencia de participación sectorial.

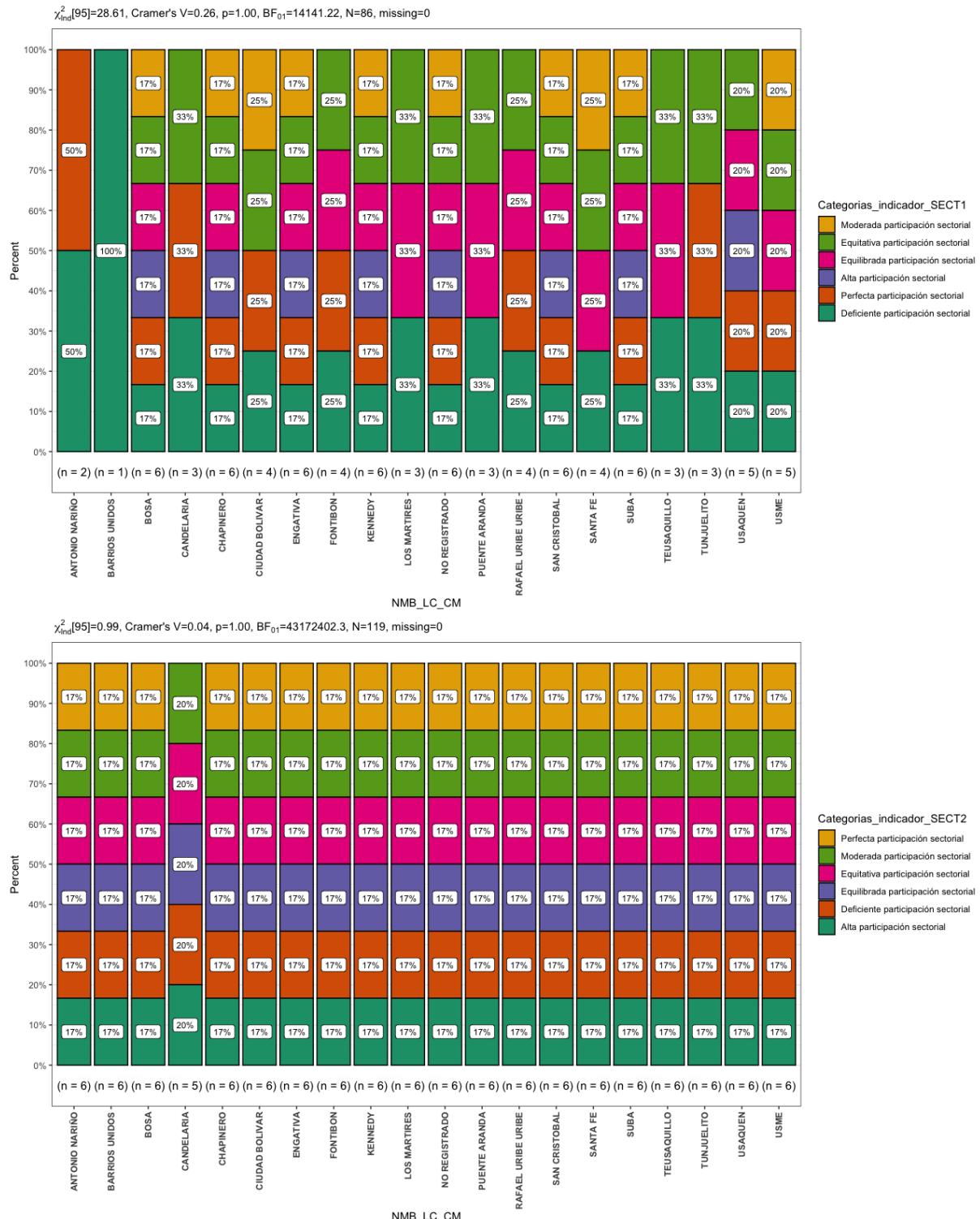
Ilustración 72. Mapa de calor de los sectores económicos.**Heatmap: sector económico primario****Heatmap: sector económico secundario**

Heatmap: sector económico terciario



En el sector económico primario se destacan las participaciones de promedio del valor de la proporción de sector económico en la categoría "Alta participación sectorial" para la localidad de "Kennedy", "Engativá", "Bosa", "San Cristobal", "Suba", "Usaquén" y "No registradas" con niveles muy similares en el intervalo del promedio que se establece para esa categoría, comparadas con las demás categorías estas resultan tener los valores más elevados, adicionalmente es posible apreciar que en la categoría de déficit de participación sectorial la localidad de "Fontibón" tiene los valores de mayor superioridad comparado con los demás. El sector económico secundario en la categoría de "Moderada participación sectorial", la localidad más representativa es "Kennedy" y para el tercer sector económico la localidad con menores niveles de participación es "Los mártires" en la categoría "Perfecta participación sectorial".

Gráfico 39. Gráfico de tablas cruzadas de los sectores económicos.



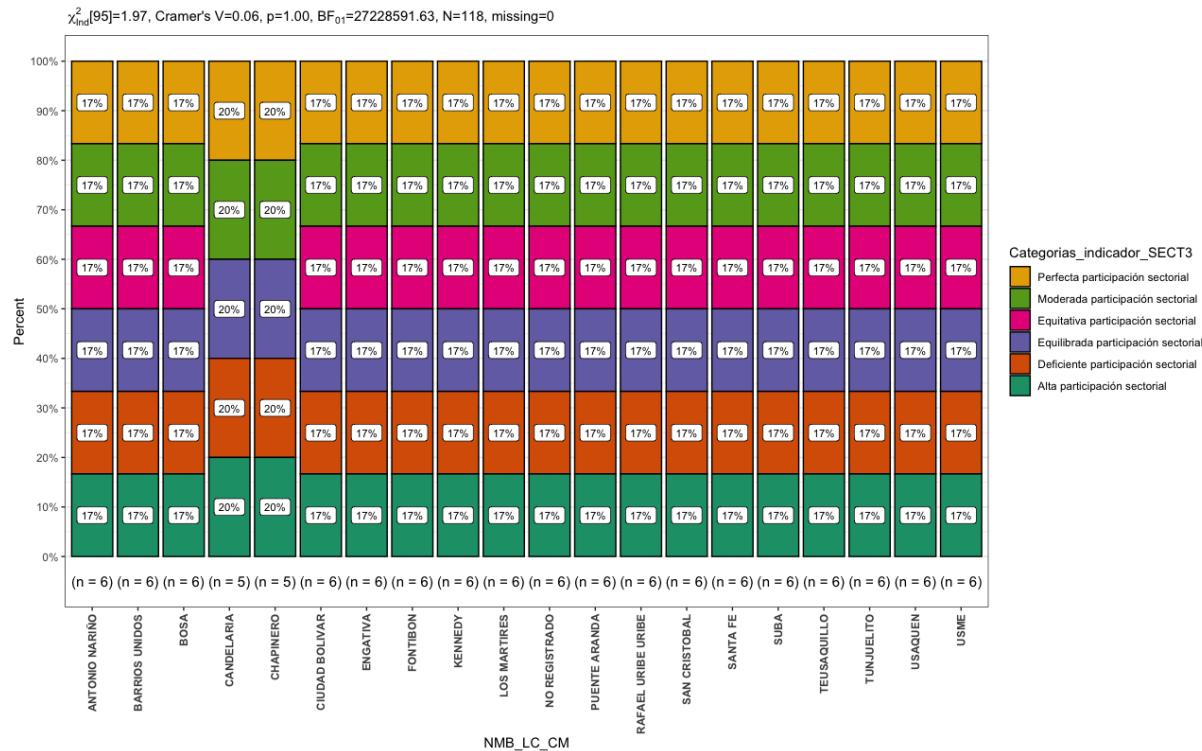
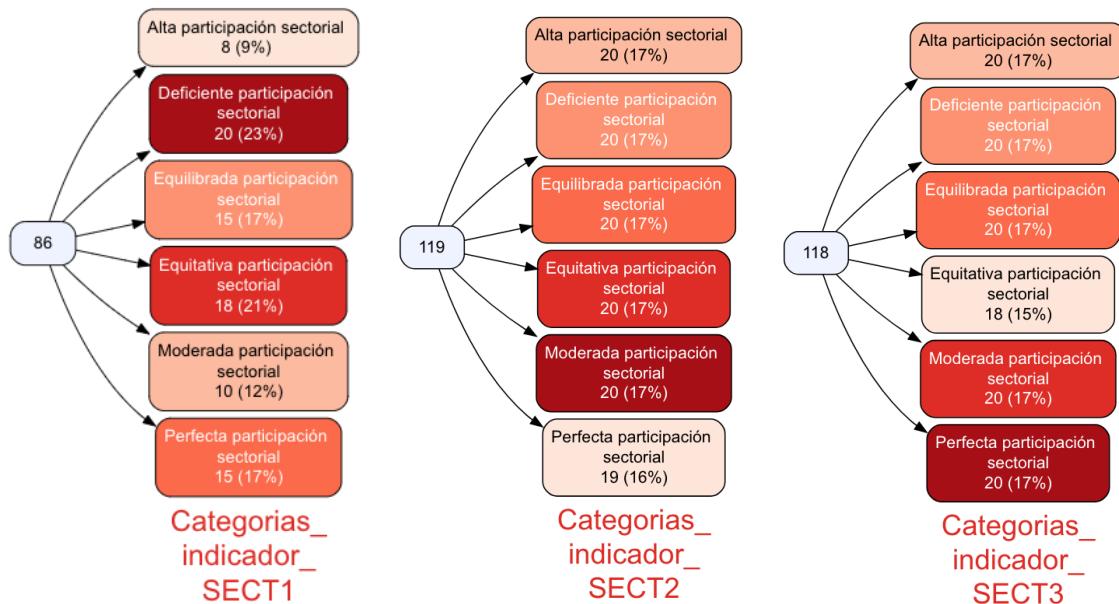


Diagrama 17. Diagrama de árbol de las tablas cruzadas para los sectores económicos.



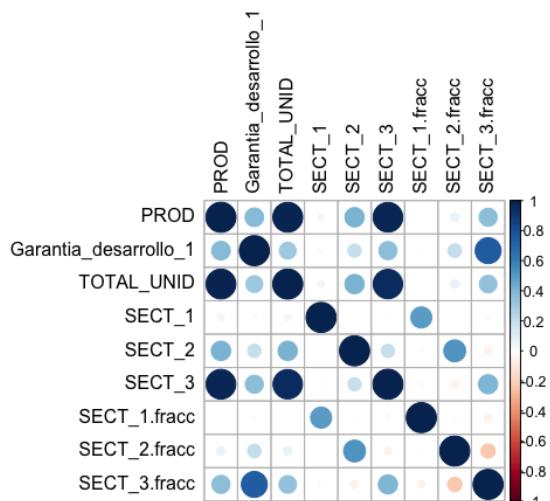
En el primer **Gráfico 40** de acuerdo a los cálculos de percentiles estimados para cada categoría en función de las localidades para el primer sector económico se presenta que el 100% de la

participación de las manzanas en estudio para la localidad de "Barrios Unidos" en la categoría "Deficiente participación sectorial", la localidad de "Antonio Nariño" refiere de un 50% de participación de manzanas en la categoría "Deficiente participación sectorial" y el 50% restante en "Perfecta participación sectorial", y las demás participaciones por categoría son posible apreciarlas en dicho primer gráfico. En el segundo gráfico, calculo estimado para el sector secundario se contempla las mismas condiciones participativas de las categorías a excepción de la localidad de la "Candelaria" que no tiene participación de manzanas de tipo Perfecta participación sectorial. Para el sector terciario de la economía se encuentran condiciones particulares para las localidades "Candelaria" y "Chapinero".

En el **Diagrama 16**, para 86 manzanas representativas en estudio el 23% se encuentran en "Deficiente participación sectorial", el 21% en "Equitativa participación sectorial", "Equilibrada participación sectorial" con 17 y el resto pueden apreciarse en el primer diagrama en el lado izquierdo. Un consolidado de 119 manzanas representativas para el sector secundario el 84% de las manzanas se encuentran con cualquiera de las categorías a excepción de la perfecta participación sectorial, la cual cuenta solo con 16% del total en estudio, a diferencia del segundo sector económico para 118 manzanas, la categoría con "Perfecta participación sectorial" asciende a 17%.

Análisis factorial para el calculo de los pesos factoriales.

Gráfico 40. Matriz de correlaciones de las variables afines al indicador GD2.



Se presenta clústeres dentro de la matriz de correlación, en las que las mayores correlaciones se muestran para las variables que representan el total de unidades productivas por manzana, el sector terciario y el indicador de garantía de desarrollo productivo.

Aplicando un test de correlación es posible conocer la incidencia de esta medida estadística con forme a los datos observados, donde :

Gráfico 41. Pruebas de Hipótesis de la medición de correlación global variables afines GD2.

$H_0: \text{variable} \rightarrow \text{Correlación} = 0$; (Indica que no existe correlación)

$H_1: \text{variable} \rightarrow \text{Correlación} \neq 0$; (Indica que si existe algún tipo de correlación)

Si H_0 es aceptada, no existe suficiente argumento para explicar AFE.

Si H_0 es rechazada, si existe suficiente argumento para aplicar AFE.

	PROD	Garantia_desarrollo_1	TOTAL_UNID	SECT_1	SECT_2	SECT_3	SECT_1.fracc	SECT_2.fracc	SECT_3.fracc	
PROD	0.00		0	0.00	0.00	0.00	0.23	0.00	0.00	0
Garantia_desarrollo_1	0.00		0	0.00	0.00	0.00	0.00	0.00	0.00	0
TOTAL_UNID	0.00		0	0.00	0.00	0.00	0.45	0.00	0.00	0
SECT_1	0.00		0	0.00	0.00	0.37	0.11	0.00	0.45	0
SECT_2	0.00		0	0.00	0.12	0.00	0.00	0.11	0.00	0
SECT_3	0.00		0	0.00	0.02	0.00	0.00	0.00	0.00	0
SECT_1.fracc	0.06		0	0.22	0.00	0.02	0.00	0.00	0.03	0
SECT_2.fracc	0.00		0	0.00	0.37	0.00	0.00	0.00	0.00	0
SECT_3.fracc	0.00		0	0.00	0.00	0.00	0.00	0.00	0.00	0

$\alpha = 0,05 \rightarrow \text{Alpha (Nivel de significancia)}$

$p - \text{value} > \alpha : \text{Se acepta } H_0$

$p - \text{value} < \alpha : \text{Se rechaza } H_0$

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Se rechaza H_0 , y se concluye que si es posible aplicar Análisis factorial, debido a que la mayoría de variables tienen un valor de p-value igual o muy próximo a cero, a excepción de las variables SECT_1.fracc y TOTAL_UNID, SECT_1 y SECT_2.fracc, no evidencian estadísticamente algún tipo de correlación, sin embargo la "Garantia_desarrollo_1" se correlaciona con todas las variables. En la fase de medición de indicadores de aplicabilidad, se utilizan distintos métodos para corroborar las condiciones actuales de los datos, y determinar con un alto grado de confiabilidad si es posible aplicar el análisis a través de lo siguiente:

Contraste de esfericidad de Bartlett

Esta técnica se emplea para determinar si a nivel de contigüidad de las variables, las correlaciones teóricas en cada par de variables es nula o no, en caso de ser nula la correlación existe un argumento suficientemente valido para aplicar el análisis factorial exploratorio.

Por tanto en esta prueba se emplea la siguiente prueba de hipótesis:

Ecuación 22. Aplicación de contraste de esfericidad de Bartlett.

H_0 : Las correlaciones teóricas entre cada par de variables es nula.

H_1 : Las correlaciones teóricas entre cada par de variables no es nulo.

$p - \text{value} > \alpha : \text{no se aplica AFE (Se acepta } H_0\text{)}$

$p - \text{value} < \alpha : \text{si se aplica AFE (Se rechaza } H_0\text{)}$

Bartlett's Test of Sphericity

```
Call: bart_spher(x = dat1)

X2 = 1862044.948
df = 36
p-value < 2.22e-16
```

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Considerando que el $p - value$ es menor que $\alpha = 0.05$, entonces se rechaza la hipótesis nula por tanto, las correlaciones teóricas entre cada par de variables es nulo, por tanto si es aplicable el AFE.

Medida de adecuaciones muestral de Kaiser, Meyer y Oklin (KMO)

Esta medida realiza un estudio variable por variable, que permite determinar si es posible aceptar la participación dentro del proceso a las variables observadas en el AFE, determinando que variables se deberían mantener, y que variables deberían ser eliminadas.

Rule of thumbs: Si KMO de una variables es mayor a 0,7 entonces se acepta que participe, de lo contrario se rechaza, respectivamente con la escala:

Tabla 39. Criterios de clasificación de la prueba de adecuación muestral KMO.

>= 0,9 : Marvelous
[0.8,0.9]: Meritorious
[0.7,0.8]: Midding(regular)
[0.6,0.7]: Mediocre
[0.5,0.7]: Miserable
<0.5 : Unacceptable

Kaiser-Meyer-Olkin factor adequacy

```
Call: KMO(r = r)
Overall MSA = 0.5
MSA for each item =
```

PROD	Garantia_desarrollo_1	TOTAL_UNID	SECT_1	SECT_2
0.5	0.5	0.5	0.5	0.5
SECT_3	SECT_1.fracc	SECT_2.fracc	SECT_3.fracc	

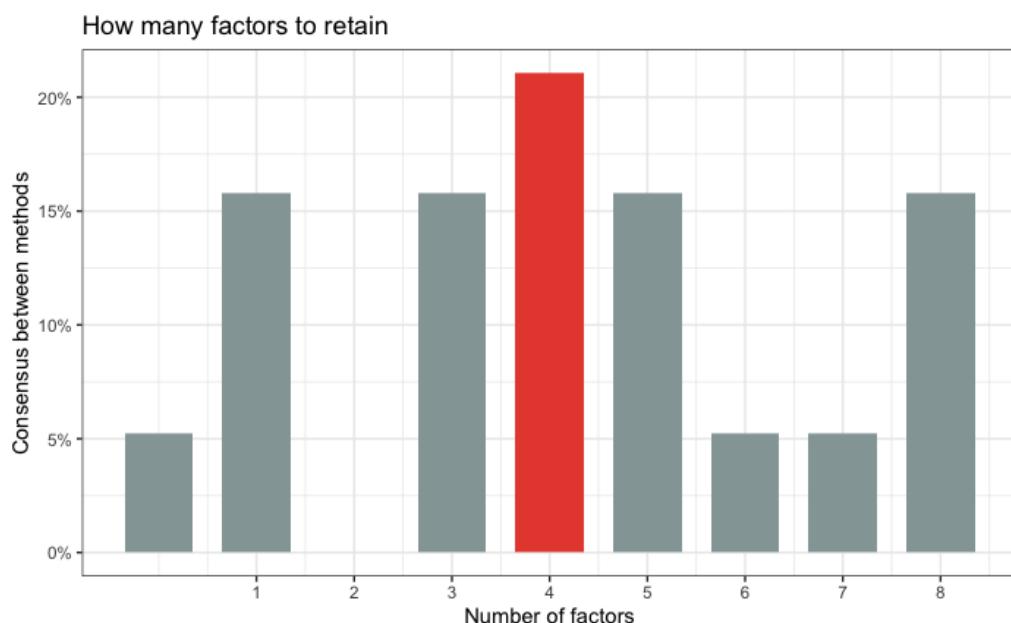
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Aplicando el test de KMO, se obtuvo que en el puntaje global es de 0.5, ubicando a todos los datos en una métrica “Miserable”, este test demuestra que las características estadísticas de las variables expresan una condición bastante precaria en cuanto a un nivel de pertinencia del uso de modelos explicativos conforme a la uniformidad en la medición.

En la etapa de implementación, se busca determinar el número de factores a extraer, a partir de diferentes métodos, no obstante también resulta importante determinar el tipo de rotación que se implementaría, el método de estimación y la matriz de estimación.

Para ello se implementa la evaluación del número de factores con una rotación de tipo "varimax", empleando la función "n_factors" para las variables en estudio y aplicando el algoritmo de máxima verosimilitud, se obtiene lo siguiente:

Gráfico 42. Gráfico de barras de consenso de métodos en función de Factores a extraer.



Este método nos indica que en el consenso entre los métodos se llega a un proporción exacta en que la cantidad de factores que debería ser extraídos es 4 (cuatro), respectivamente:

Ilustración 73. Reporte del consenso entre métodos AFE.

Method Agreement Procedure:

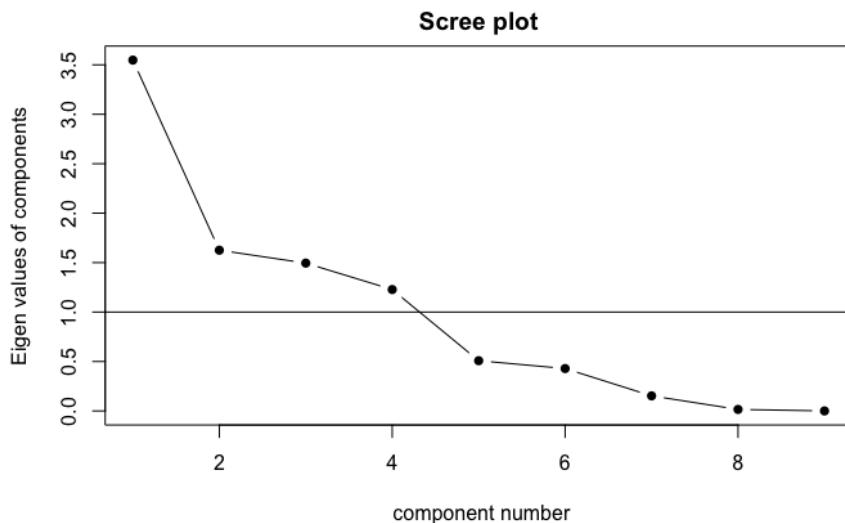
The choice of 4 dimensions is supported by 4 (21.05%) methods out of 19
(beta, Parallel analysis, Kaiser criterion, Scree (SE)).

Los resultados plantean el 21.05% es decir 4 métodos apoyan la idea que sean seleccionados 4 factores dimensionales, estos 4 métodos de análisis factorial serían el beta, Parallel Analysis, el Kaiser criterion y el Scree(SE), como se presentan a continuación:

Tabla 40. Métodos disponible (Total :19) para el Análisis Factorial Exploratorio.

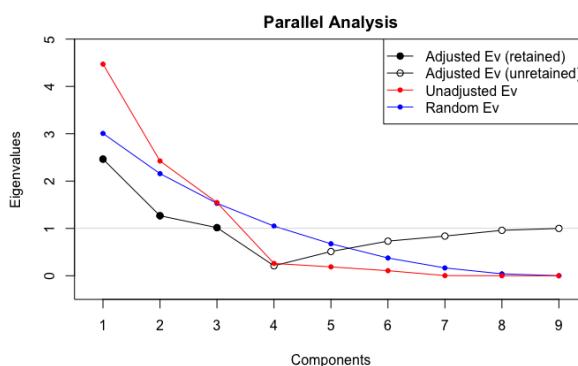
n_Factors	Method	Family
<dbl>	<chr>	<chr>
0	Scree (R2)	Scree_SE
1	Optimal coordinates	Scree
1	Acceleration factor	Scree
1	Velicer's MAP	Velicers_MAP
3	CNG	CNG
3	t	Multiple_regression
3	p	Multiple_regression
4	beta	Multiple_regression
4	Parallel analysis	Scree
4	Kaiser criterion	Scree
4	Scree (SE)	Scree_SE
5	VSS complexity 1	VSS
5	BIC	BIC
5	BIC (adjusted)	BIC
6	Bentler	Bentler
7	VSS complexity 2	VSS
8	Bartlett	Barlett
8	Anderson	Barlett
8	Lawley	Barlett

Otro método para determinar el número factores a extraer, es el destacado a partir del cálculo de los eigenvalues como resultado de la matriz de correlaciones, este método realiza una ponderación de acuerdo a las correlaciones registradas y estima según los pesos de cada variable las mejores condiciones de las componentes que podrían explicar las variables entre sí, se presenta a continuación el resultado de su implementación:

Gráfico 43. Gráfico método Scree AFE.

El diagrama de dispersión es una representación visual de los valores propios. La inspección visual del diagrama de dispersión es una forma rápida y sencilla de hacerse una idea de la dimensionalidad de su conjunto de datos, los valores que estén por encima de el peso de 1, son considerados óptimos, y las cantidades de componentes contados desde esa medición serían el número de factores a elegir, el resultado que el número de factores que deberían ser elegidos es de 4.

En el método de Parallel, se determino que el comportamiento de los componentes o factores a seleccionar es el mismo implementando la función metodológica de "PA(Principal Axes)", y los componentes axiales reflejan de una simulación con una proyección no tan ajustada debido a que los valores del eigenvalues son mejor explicados de acuerdo con los valores actuales.

Gráfico 44. Gráfico de análisis Parallel.

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

Análisis factorial exploratorio:

En el proceso de explicación con los factores seleccionados, en este caso la selección es para el método que se integran parámetros como numero de factores igual a 4 y una rotación de tipo "oblimin".

De acuerdo a las variables mostradas en el resultado de la aplicación de EFA. A continuación se explican las métricas factoriales:

+PC1: Cargas factoriales de cada variable.

+ h^2 : Comunalidad(Varianza común explicada)

+ u^2 : Especificidad (Varianza residual) o varianza no explicada.

Para el caso de los datos, se obtiene el siguiente resultado:

Tabla 41. Resultado de AFE : Cargas factoriales.

	MR1 <S3: AsIs>	MR3 <S3: AsIs>	MR2 <S3: AsIs>	MR4 <S3: AsIs>	h^2 <dbl>	u^2 <dbl>	com <dbl>
PROD	0.99	0.02	0.07	0.01	1.0041021	-0.0041020842	1.011709
Garantia_desarrollo_1	-0.02	0.99	0.15	0.03	0.9994497	0.0005503377	1.050026
TOTAL_UNID	1.00	-0.03	0.07	0.01	0.9952670	0.0047329549	1.012135
SECT_1	0.04	0.00	-0.01	0.57	0.3208071	0.6791928668	1.011298
SECT_2	0.30	-0.01	0.63	-0.01	0.5177751	0.4822249483	1.433040
SECT_3	0.96	0.04	-0.14	-0.02	0.9531698	0.0468301914	1.044917
SECT_1.fracc	0.00	0.01	-0.02	0.87	0.7580673	0.2419327147	1.001362
SECT_2.fracc	-0.06	0.06	0.83	-0.02	0.6962741	0.3037259162	1.022090
SECT_3.fracc	0.12	0.76	-0.31	-0.05	0.7434495	0.2565504790	1.383167

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

** A mayor h^2 , mejor será el modelo 0;1

**A menor u^2 , mejor será el modelo 0;1

Conclusión:

+La variable "PROD", es explicada en un 100%,seguido de "Garantia_desarrollo_1" es explicada en un 99,94%, "TOTAL_UNID" es explicada al 99.52%.

+En promedio las variables SECT_1,SECT_2 y SECT_3 son explicadas en un 59.72% y las variables SECT_1.fracc, SECT_2.fracc y SECT_3.fracc son explicadas en un 73.25%, por lo tanto estas variables fraccionadas pierden menor información.

Ilustración 74. Reporte de resultados del AFE.

```

Factor Analysis using method = minres
Call: fa(r = dat1, nfactors = 4, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix

          MR1  MR3  MR2  MR4
SS loadings  3.05 1.60 1.25 1.08
Proportion Var 0.34 0.18 0.14 0.12
Cumulative Var 0.34 0.52 0.66 0.78
Proportion Explained 0.44 0.23 0.18 0.15
Cumulative Proportion 0.44 0.67 0.85 1.00

With factor correlations of
      MR1  MR3  MR2  MR4
MR1  1.00 0.36 0.08 -0.01
MR3  0.36 1.00 0.03 -0.02
MR2  0.08 0.03 1.00 0.03
MR4 -0.01 -0.02 0.03 1.00

The root mean square of the residuals (RMSR) is 0.02
The df corrected root mean square of the residuals is 0.04

The harmonic number of observations is 43408 with the empirical chi square
712.48 with prob < 1.2e-150
The total number of observations was 43408 with Likelihood Chi Square =
11175537 with prob < 0

Tucker Lewis Index of factoring reliability = -53.684
RMSEA index = 6.55 and the 90 % confidence intervals are 6.541 NA
BIC = 11175473
Fit based upon off diagonal values = 1

```

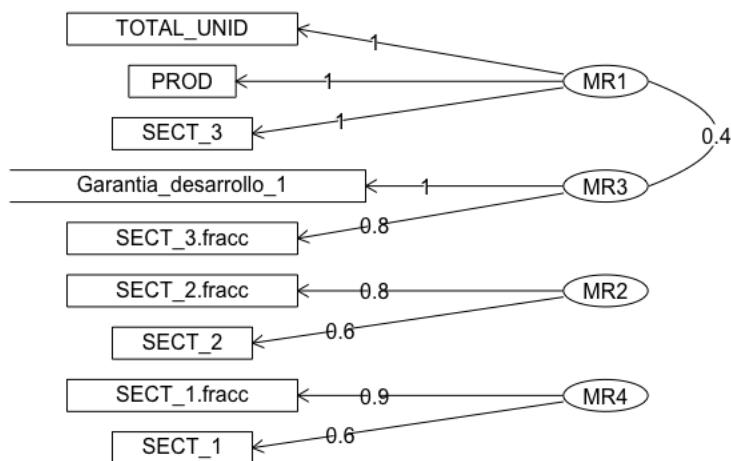
Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

**Lo ideal es que la "Proportion Var" acumulada sea lo más cercano a 1.

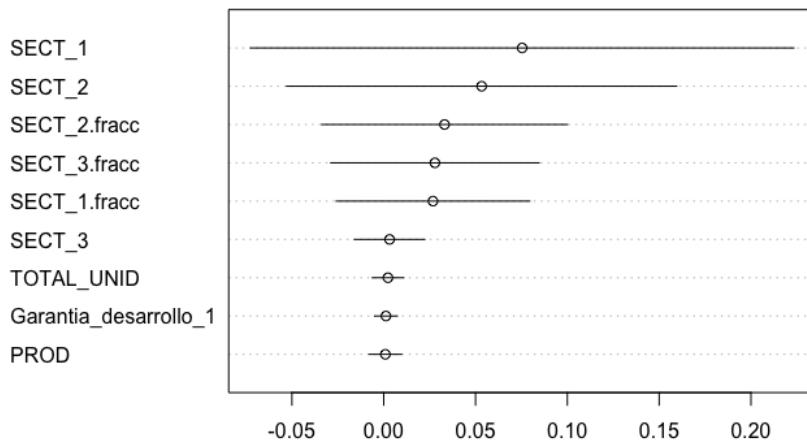
+The root mean square of the residuals (RMSR) is 0.04, correspondiente a la raíz cuadrada media de los residuos, esta mide teóricamente los residuos modelo AFE. Es un buen indicador cuando su valor es menor que 0.08, en este caso presenta excelente resultado, y la varianza explicada acumulada para todos los factores es de 78%, no se podrían explicar en su totalidad pero es posible generar planteamientos a partir de ello expresa una alta aceptación. Los resultados obtenidos, se explican a partir de los gráficos y ecuaciones factoriales. A continuación se presentan los diagramas explicativos del análisis factorial para el método seleccionado.

Diagrama 18. Análisis factorial exploratorio gráfico.

Análisis factorial confirmatorio para el indicador Garantía de desarrollo por sectores



Confidence Intervals around the mean



Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

En este primer diagrama es posible apreciar los factores seleccionados y la relación ponderada explicativa de estos factores con respecto a las variables observadas, es de observarse que para el Tercer factor (MR3), estima que debería relacionar la variable del tercer sector económico fraccionada o en proporción con el indicador "Garantía de desarrollo productivo" debido que en promedio los explica en un 90%, lo que indica que ambas variables tienen características

que podrían compartir, este factor se correlaciona con el Factor Número 1 (MR1) en una escala de 0.4, el cual comparte directamente una explicación del 100% de la información para las variables "TOTAL_UNID", "PROD" y "SECT_3".

En ese orden de idea, los datos presentados al modelo AFE, consideran en su ajuste relativo conforme a las condiciones estadísticas de cada variable y establece las métricas para las variables en un orden de importancia, por tanto la variable más importante para el factor MR3 es "Garantia_desarrollo_1", dándole continuidad al contexto, podría decirse que para el factor MR1 todas las variables en la que son parte de su entorno explicativo, refieren de la misma importancia relativa y que para cada factor diferentes de los mencionados las variables más importantes son las fraccionadas. Las variables que hacen parte del primer factor tienen el intervalo de confianza alrededor de la media de residuos con valores muy pequeños y que los valores de los sectores económicos fraccionados o por participación tienen los intervalos que contienen valores positivos, esto permitiendo inferir que no se puede descartar la posibilidad de la media de los residuos en el calculo de pesos factoriales y puntajes en promedio sea nula o estime a valores muy pequeños.

Ilustración 75.Calcuilo de las cargas factoriales por cada variable (Agrupación en términos de peso).

Loadings:

	MR1	MR3	MR2	MR4
PROD	0.985			
Garantia_desarrollo_1		0.992	0.153	
TOTAL_UNID	1.002			
SECT_1			0.566	
SECT_2	0.301		0.633	
SECT_3	0.962		-0.138	
SECT_1.fracc				0.871
SECT_2.fracc			0.835	
SECT_3.fracc	0.117	0.763	-0.309	

Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado de RStudio, Observatorio Inmobiliario Nacional DANE.

En las cargas de cada variable como resultado del análisis factorial exploratorio, es posible contemplar que para cada factor la puntuación de carga factorial es diferente y los valores indican que tanta pertinencia tendrá un factor y en que magnitud podría pesar estos valores varía en las dimensiones de los números reales, en las cuales los valores negativos expresan que la variable no supondría ningún tipo de explicación sino que por el contrario perturbaría la

integración de las variables cuyo peso en dicho factor explicativo esta siendo explicada en más de un 0.7 peso factorial.

Para determinar el indicador representativo de garantía de desarrollo por sector económico ponderado, se consideraran las máximas cargas factoriales registradas por cada variables orgánica de "SECT" (referente a sectores), debido a que en ningún factor estas tres variables se organizan con cargas distribuidas, por tanto las ponderaciones a considerar serian:

Ilustración 76. Pesos factoriales pre-seleccionados para la construcción del GD2.

Loadings:

	MR1	MR3	MR2	MR4
SECT_1				0.566
SECT_2		0.301		0.633
SECT_3		0.962		-0.138

Los máximos pesos factoriales registrados son: Para la variable "SECT_1" es de 0.566, "SECT_2" es de 0.633 y para el "SECT_3" es de 0.962. A continuación se precede a calcular las ponderaciones utilizadas para calculo del estimador de garantía de desarrollo por sector económico ponderado, esto es implementando la normalización a través el calculo de proporciones por cada peso factorial, en primera estancia se calcula el total de peso factorial observado:

$$\text{Peso factorial total (PFT)} = PF1 + PF2 + PF3$$

Donde PF1, corresponde al máximo peso factorial del sector 1, PF2 el máximo peso factorial del sector 2 y PF3 el máximo peso factorial registrado para el sector 3.

Una vez calculado, se estima entonces la proporción de ese peso factorial conforme al total

$$\begin{aligned} \text{Normalización sector 1 (NS1)} &= \frac{PF1}{PFT} \\ \text{Normalización sector 2 (NS2)} &= \frac{PF2}{PFT} \\ \text{Normalización sector 3 (NS3)} &= \frac{PF3}{PFT} \end{aligned}$$

Por tanto, esta normalización permite conocer las ponderaciones que permitirán estimar el indicador de garantía de desarrollo por sector económico ponderado, con la siguiente ecuación:

$$\text{Garantía}_{\text{desarrollo}}2 = NS1 * SECT_1.\text{frac} + NS2 * SECT_2.\text{frac} + NS3 * SECT_3.\text{frac}$$

El resumen estadístico resultante es el siguiente:

Ilustración 77. Resumen estadístico con estimación del indicador "Garantía de desarrollo por sector económico".

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1 Garantia_desarrollo_prod	0	1	0.544	0.636	-1	0	1	1	1	
2 SECT_1.fracc	0	1	0.00335	0.0490	0	0	0	0	1	
3 SECT_2.fracc	0	1	0.0991	0.232	0	0	0	0.0240	1	
4 SECT_3.fracc	0	1	0.596	0.452	0	0	0.867	1	1	
5 Garantia_desarrollo_sect	0	1	0.663	0.443	0	0	0.954	1	1	

Teniendo en cuenta que los valores empleados fueron los siguientes:

Tabla 42. Valores de las normalizaciones para cada sector económico (ponderación o peso)

Normalización	Valor
NS1	0.2619
NS2	0.2929
NS3	0.4451

0.2619158

0.2929199

0.4451643

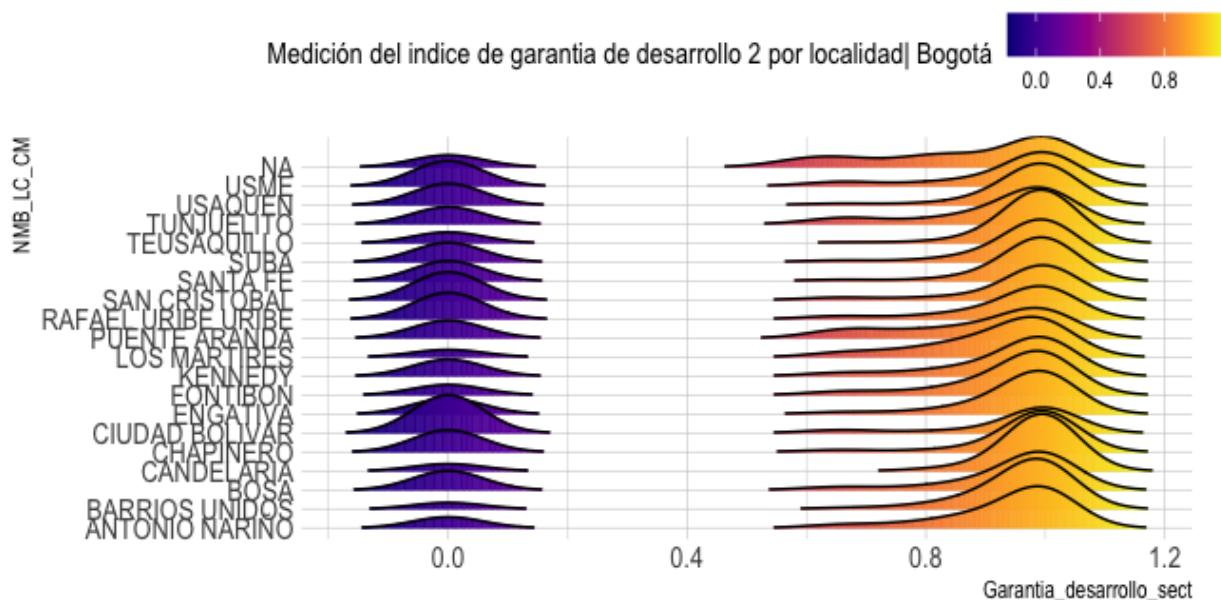
En el resumen de la **Ilustración 77**, es de notarse que de los histogramas representados para cada sector económico, el sector terciario es el que más similitud tiene con el indicador estimado de GD2, es de notarse que los datos se encuentran en una escala del 0 al 1, debido a que una vez se procedió a las operaciones fue entonces rescalado. En la normalización contenida en la tabla el sector con mayor ponderación resulta ser el sector de servicios el cual según el análisis factorial es el que tiene mejor puntaje y relacionamiento con las demás variables.

Análisis categórico exploratorio del indicador: "Garantía de desarrollo 2" (GD2).

El gráfico de densidad del indicador, es de apreciarse que el volumen de datos que se concentran en el valor 1 del indicador es mayor que los que se concentran en el valor 0 (cero) y -1, comparado para la localidad de "Ciudad Bolívar", esta presenta mayor cuantía de valores concentrados en 0 (cero) que en ambos extremos del indicador, esto indica que en esa manzana existen nulidad de garantía de desarrollo.

Gráfico 45. Gráfico de densidad del indicador GD2 por localidad

MGN- Nivel Manzana :Dataset 2021



Para entender el comportamiento de los datos de este indicador, se procede a clasificar los valores del indicador en diferentes categorías, presentadas a continuación:

$[-\infty, 0.15] \rightarrow \text{"Garantía Nula"}$

$[0.15, 0.58] \rightarrow \text{"Deficiente garantía"}$

$[0.58, 0.59] \rightarrow \text{"Perfecta garantía SECTOR 1"}$

$[0.59, 0.66] \rightarrow \text{"Perfecta garantía SECTOR 2"}$

$[0.66, 0.99] \rightarrow \text{"Garantía Politómica"}$

$[0.99, 1] \rightarrow \text{"Perfecta garantía SECTOR 3"}$

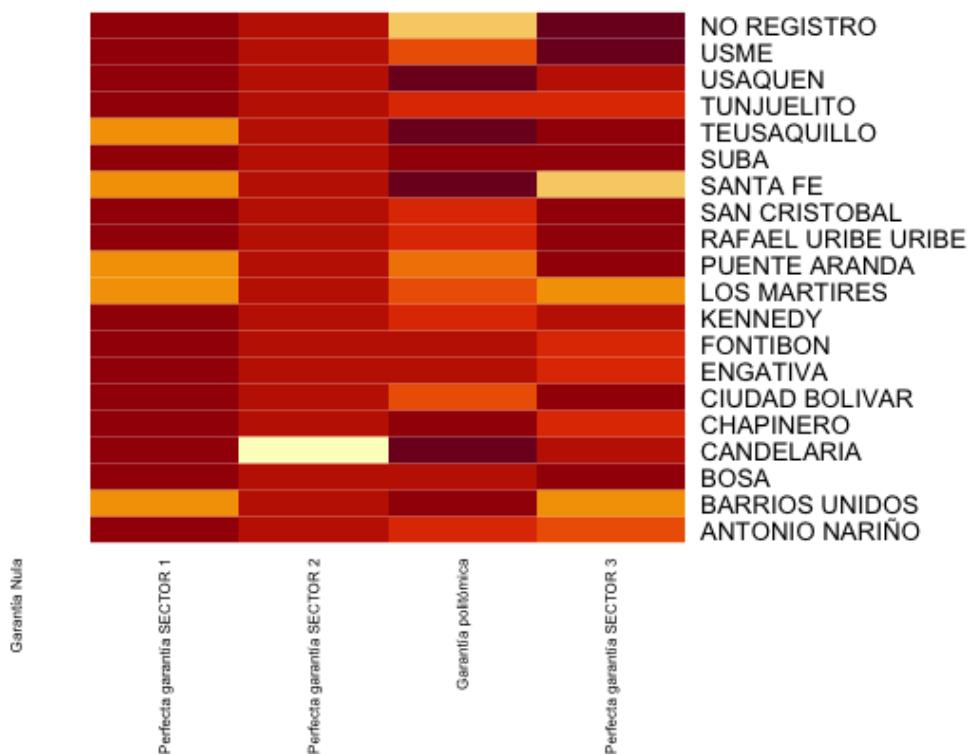
A continuación se presenta el resumen estadístico de estas categorías:

Tabla 43. Resumen de categorías indicador Garantía de desarrollo 2.

Garantía Nula	Deficiente garantía	Perfecta garantía	SECTOR 1	Perfecta garantía	SECTOR 2
13082	0		80		1609
Garantía politómica	Perfecta garantía	SECTOR 3			
9341	19296				

Como es de apreciarse, el 44.45% de los datos se concentran en la categoría "Perfecta garantía SECTOR 3", el 21.52% concentrados en la "Garantía polítómica" esto refiere a la idea que en esas manzanas existen unidades productivas con más de 2 sectores, un 30.14% con "Garantía Nula" lo que refiere a la idea que en esas manzanas no existen unidades productivas, así acumulando un 96,30% en estas 3 categorías, el 13.7% restante corresponde para las manzanas con perfecta garantía en el sector 2.

Ilustración 78. Mapa de calor del indicador Garantía desarrollo 2.



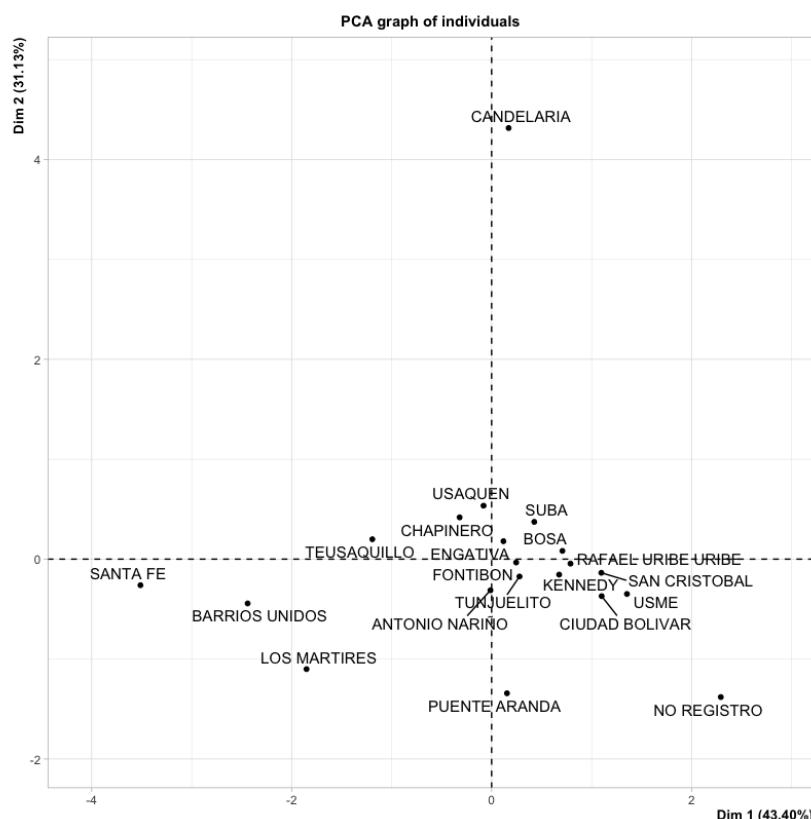
En la categoría "Garantía Politómica", se presentan valores extremos para las localidades tales como "Candelaria", "Santa Fé", "Teusaquillo" y "Usaquén", lo que indica que en estos lugares se encuentran manzanas que ocupan mayor porcentaje de participación por unidades en el sector 3 que son parcialmente integradas por alguna de los dos otros sectores económicos en bajo estándares. En las categorías, donde se discriminan los datos que ocupan una perfecta garantía

de desarrollo sea para el sector 1 o para el sector 2, es de notarse que las medidas del indicador resultan ser uniformes en estas categorías, excepción de "Teusaquillo", "Santa fé", "Puente Aranda", "Los Martires" y "Barrios Unidos", lo cuales para el indicador, presentan mayor ausencia del sector primario, efecto que es de encontrarse a su vez en la categoría "Perfecta garantía SECTOR 2", en la cual la localidad "Candelaria" demuestra tener inexistencia de manzanas que se ocupen con unidades alusivas a la producción industrial. De acuerdo al indicador, en la categoría "Perfecta garantía SECTOR 3", la localidad con mayor registros en promedio con valores de este indicador altos, entendiéndose que cuanto mayor o cercano a uno indica que en esa manzana solo se dedican a actividades de tipo servicios, la localidad de "USME" es la que más alto presente este índice, por tanto es mucho más probable encontrar actividades asociadas a las instituciones, comerciales y de servicios.

En el análisis PCA, dos componentes explican 74.53% de la información de los datos, respectivamente:

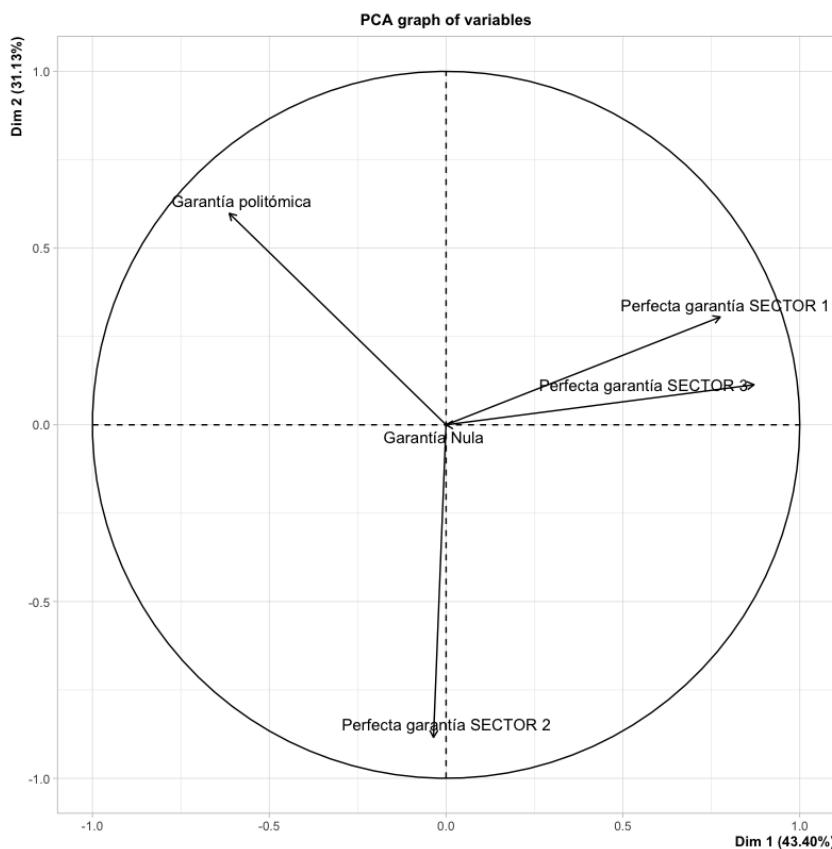
Aplicando el método espacial individual se obtiene:

Diagrama 19. PCA Método Gráfico Espacial individual por localidades : GD2



Para las localidades que representan el peso de participación de cada localidad en función del indicador del garantía de desarrollo, existe presencia de localidades que se encuentran concentradas en los cuadrantes inferiores en el método gráfico del PCA, es de notarse que la localidad de "Candelaria" se encuentra altamente alejada de las demás localidades, esto indica que el perfil de dicha localidad no se ajustan a ninguno en particular y que su dirección en términos de mediciones promedio del indicador refiere de situaciones particulares. Por su parte, las localidades en el 4to cuadrante (inferior derecho), de acuerdo al método grafico no presentan tanta dispersión, sino por el contrario existen mayor presencia de localidades que se agrupan, estas tienen perfiles similares en un mismo cuadrante, indicando alta correlación entre estas localidades.

Diagrama 20. PCA método circular.



En el método circular, la categoría "Garantía Politómica" va en diferente dirección que las otras categorías, en las categorías "Perfecta garantía SECTOR 1" y "Perfecta garantía SECTOR 3" van dirección con poca diferencia entre ellas (baja variación en el ángulo de inclinación), el mismo sentido y magnitud, es decir que para las localidades que se encuentren en estos cuadrantes cuentan con la característica de tener la tendencia que en sus manzanas se encuentren

perfectamente articuladas con unidades productivas en estos dos sectores económicos ; en la categoría "Garantía polítómica", presenta significativas diferencias en la composición de los vectores de datos en las tablas cruzadas y debido a que esta se encuentra en un solo cuadrante no se correlaciona con ninguna otra categoría, tal como le resulta a la categoría "Perfecta garantía SECTOR 2", contrastando estos planteamiento de referencia conforme a las direcciones de cada categoría en el método PCA circular, es de notarse que la localidad de "Santa Fé" se encuentra en la mitad de las apreciaciones direccionales de las categorías "Garantía polítómica" y "Perfecta Garantía SECTOR 2", esto soportando la idea que esta localidad tiene mayor participación de manzanas instruidas con unidades productivas en el sector 2, a que se encuentran más de 2 sectores.

Gráfico 46. Gráfico de tablas cruzadas del indicador Garantía de desarrollo 2.

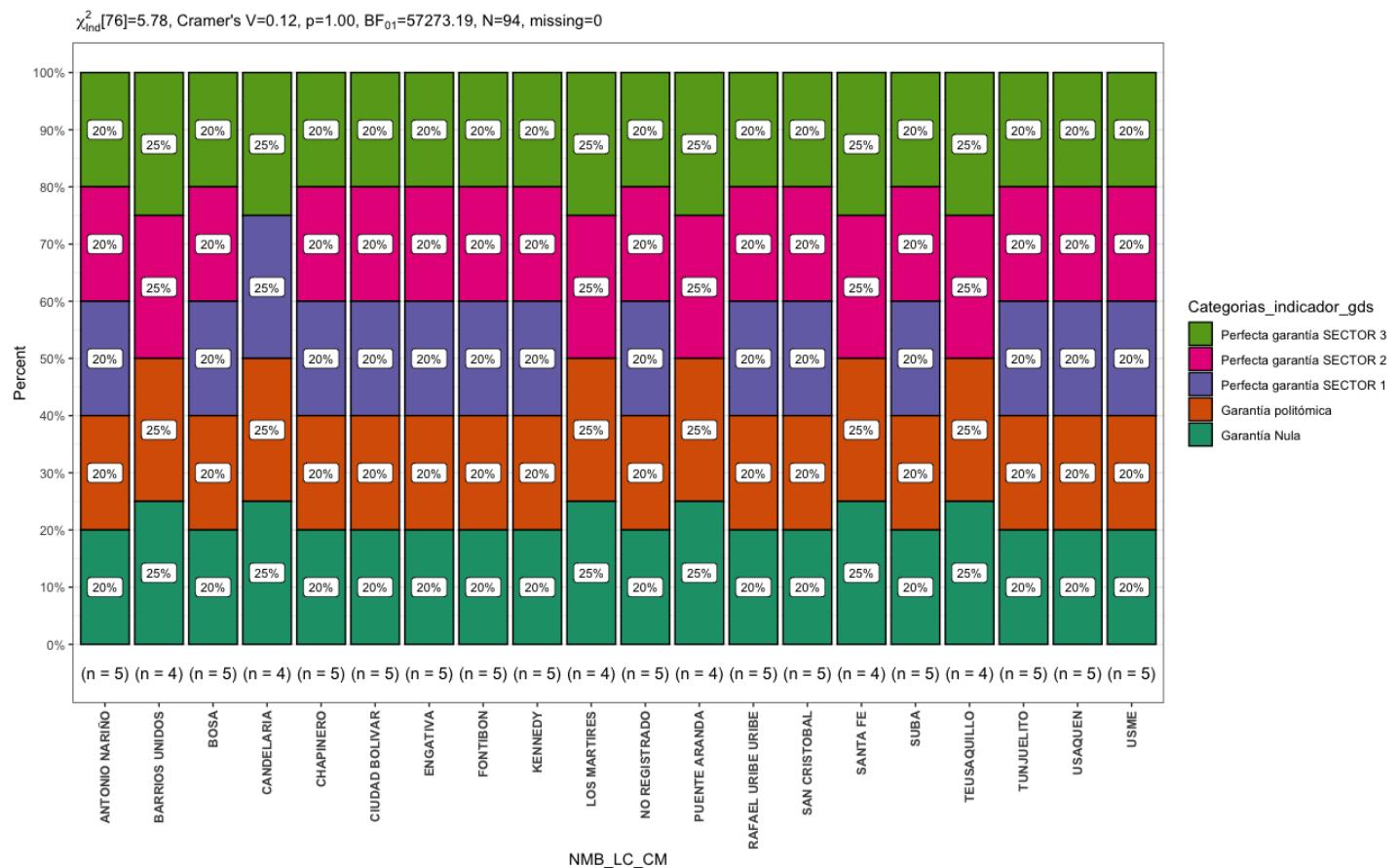
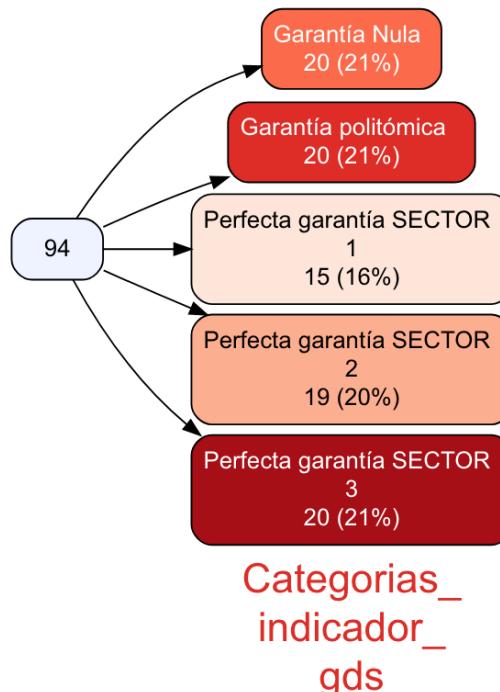


Diagrama 21. Diagrama de árbol de las tablas cruzadas en el indicador Garantía de desarrollo 2.



En el **Gráfico 46**, se encontró que 14 localidades cuentan con igualdad de participación para las 5 (cinco) categorías, estas refieren que las localidades en sus porciones de manzanas reflejan un intervalo de confianza en función de la proporción de las categorías mencionadas, por tanto es normal que dentro de estas localidades 1 de cada 5 manzanas se encuentren en cualquiera de las categorías. Las localidades como "Barrios Unidos", "Los Martires", "Puente Aranda", "Teusaquillo", "Santa fé" y "Barrios Unidos" excluyen la categoría de "Perfecta garantía del SECTOR 1" y estas tienen participación del 25% en cada categoría, asegurándose que al menos 75% de las manzanas contengan alguna unidad productiva en los sectores económicos. Las localidades con mayor balance de categorías productivas son las de "Puente Aranda" y "Antonio Nariño" debido a que el 75% de las manzanas se encuentran en las mejores categorías de garantía de desarrollo. La localidad de "Candelaria" muestra una exclusión de la participación en la categoría "Perfecta garantía SECTOR 2", concentrándose que 1 de cada 4 manzanas en esta localidad refiere de cualquier categoría a excepción del sector secundaria de la economía, o se puede considerar despreciable su participación.

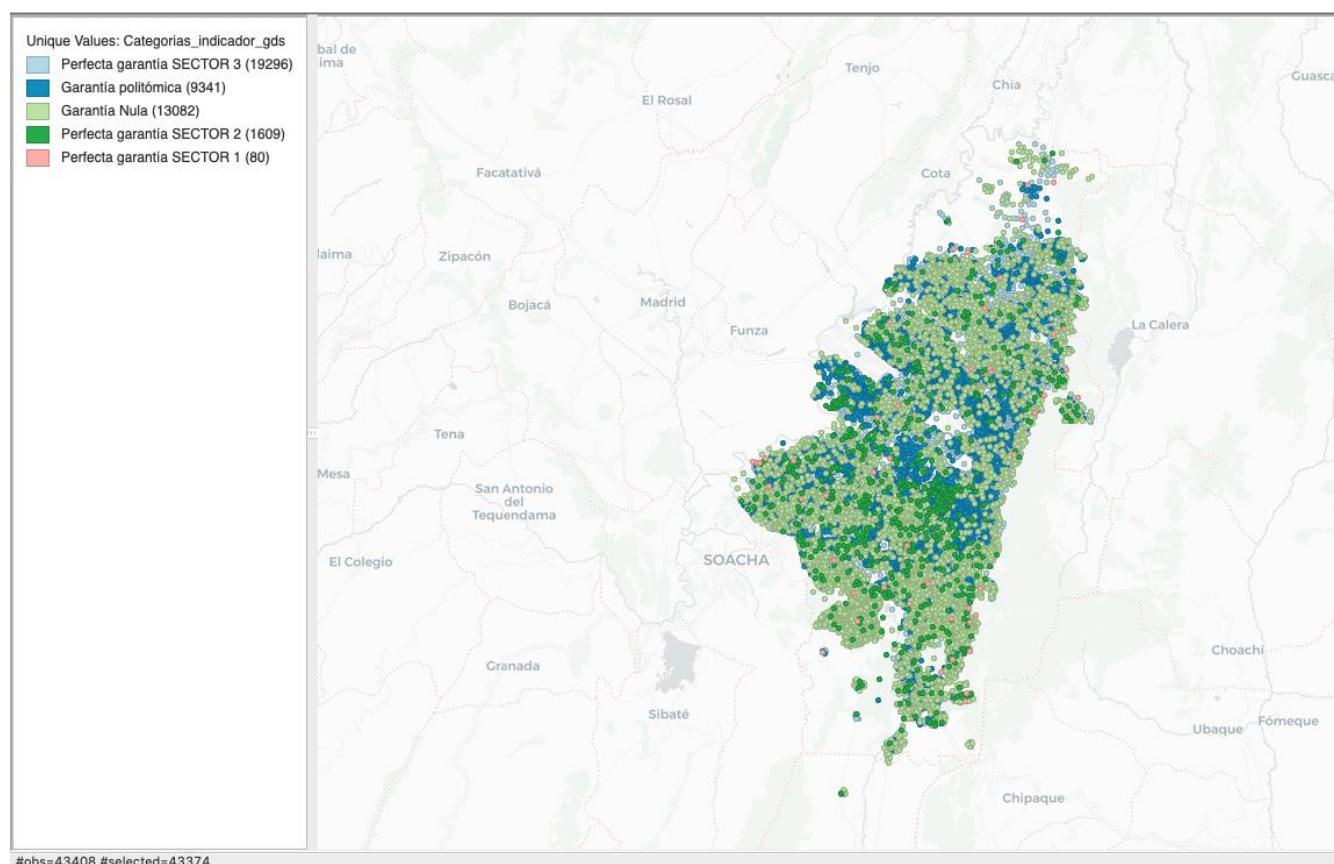
En el **Diagrama 20**, se obtuvo para el calculo de 94 manzanas representativas que agrupan los promedios para cada categoría, diferentes participaciones. Para la categorías "Garantía nula", "Garantía política" y "Perfecta garantía SECTOR 3" con 21% para cada una, "Perfecta garantía

SECTOR 2" con un 20% acumulando un total del 84% de las manzanas representativas, el complemento de esta participación se agrupa para las manzanas representativas con un 16% para la categoría "Perfecta garantía SECTOR 1".

Análisis espacial de los resultado de la medición del indicador “Garantía de desarrollo por sector económico”:

Mapas temáticos del indicador :GD2.

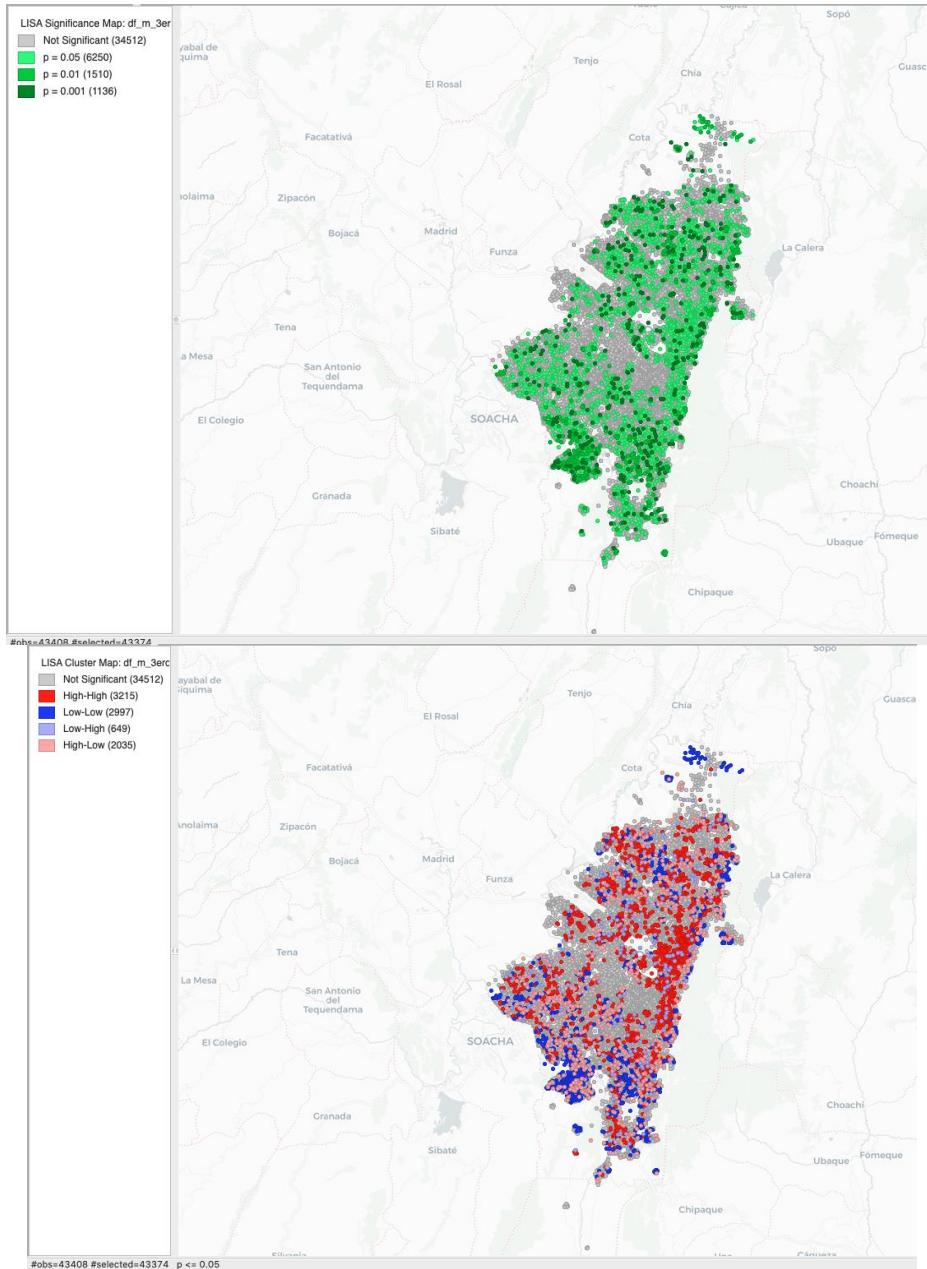
Ilustración 79. Mapa de distribución geoespacial de las categorías del índice GD2



En la **Ilustración 79**, se presenta un mapa único de valores del indicador en que se puede apreciar que para la categoría "Perfecta garantía SECTOR 1", se registran solo 80 manzanas y que estas están distribuidas en toda la ciudad con mucha dispersión, en cambio las zonas que no se consideran productivas debido a que no existe garantía de actividades productivas en esas manzanas se ubican en la garantía "Garantía Nula" de forma más uniforme en toda la ciudad con 13082 manzanas.

Análisis de auto correlación espacial del indicador:

Ilustración 80. Mapas de significancia y Clúster : Método de “Índice de Morán y LISA” 95% confiabilidad y 999 permutaciones.



Se observa que para las diferentes entidades de la capa que están formando algún tipo de agrupación espacial de tipo clústeres de valores en la categoría de Baja garantía de

productividad, el nivel del valor para el indicador garantía de desarrollo de significación estadística es alto de una forma aleatoria en la ciudad. La probabilidad de encontrar manzanas con valor correlacionadas con la entidad de contigüidad próxima o los vecinos próximas y viciosa, o agrupaciones de manzanas que comparten características bajas en significancias y que ambas estén sujetas a condiciones significativas iguales de mediciones del indicador de valores bajo se muestran con color "azules" (Bajos y Bajos) mostrando grado de significancia igual que el 0,05, existe un alta presencia de zonas de color gris en el mapa de significancia, esto quiere decir que no existe pesos significativos en como esta distribuido el indicador como argumento suficiente para no crear agrupaciones en zonas en las que se pueda evidenciar altos niveles de garantía de desarrollo rodeado de zonas con estas mismas características (en la leyenda presenta que el "High-High", es decir esas zonas ubicadas con agrupaciones de manzanas con alta garantía de productividad) es inexistentes al momento de permutar 999 los valores del peso de contigüidad (vecindad) en el 95% de los casos. En el 95% de los casos es normal encontrar manzanas cuya participación en sector económico terciario tenga una relación del 100%, es decir el total de unidades productivas en esa manzanas sea equivalente al total de unidades productivas que garantizar realizar actividades productivas de servicio, comercio y/o institucional que favorece ese segmento económico en los vectores de datos presentados con color rojo en la Ilustración

Ilustración 81. Mapas comparativos de significancia estadística con un 95% de confiabilidad HotSpots

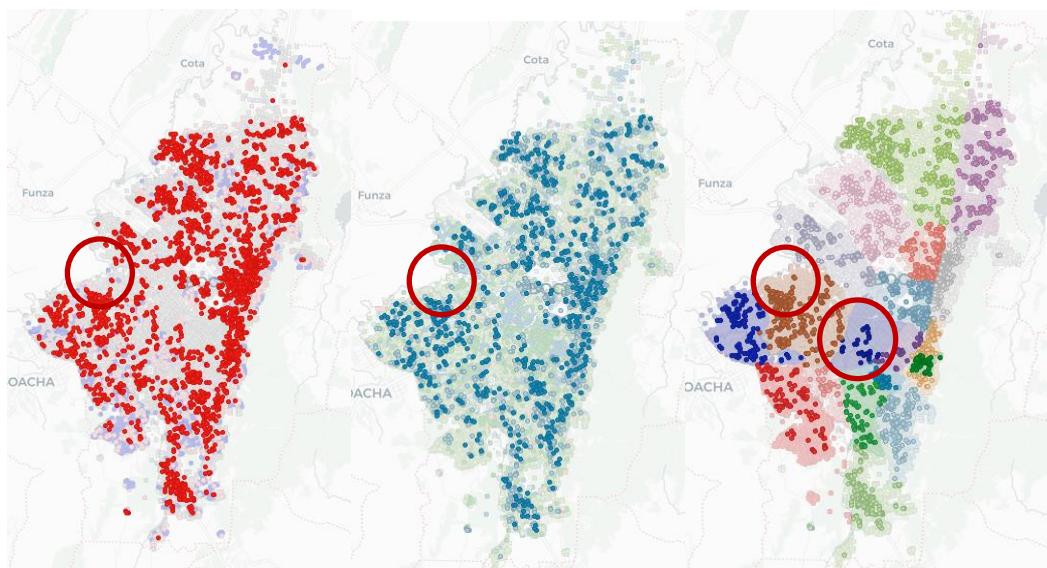
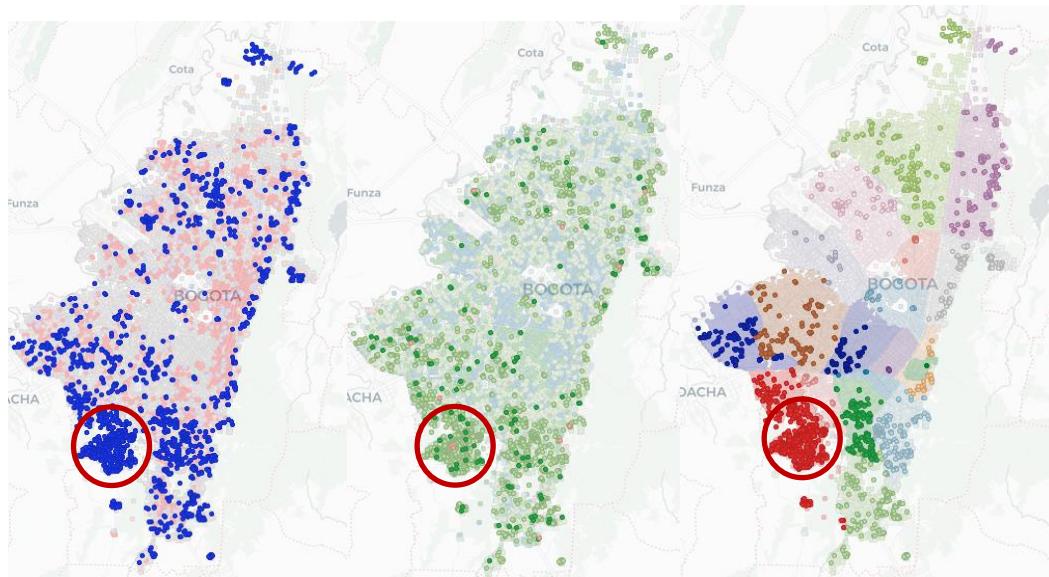
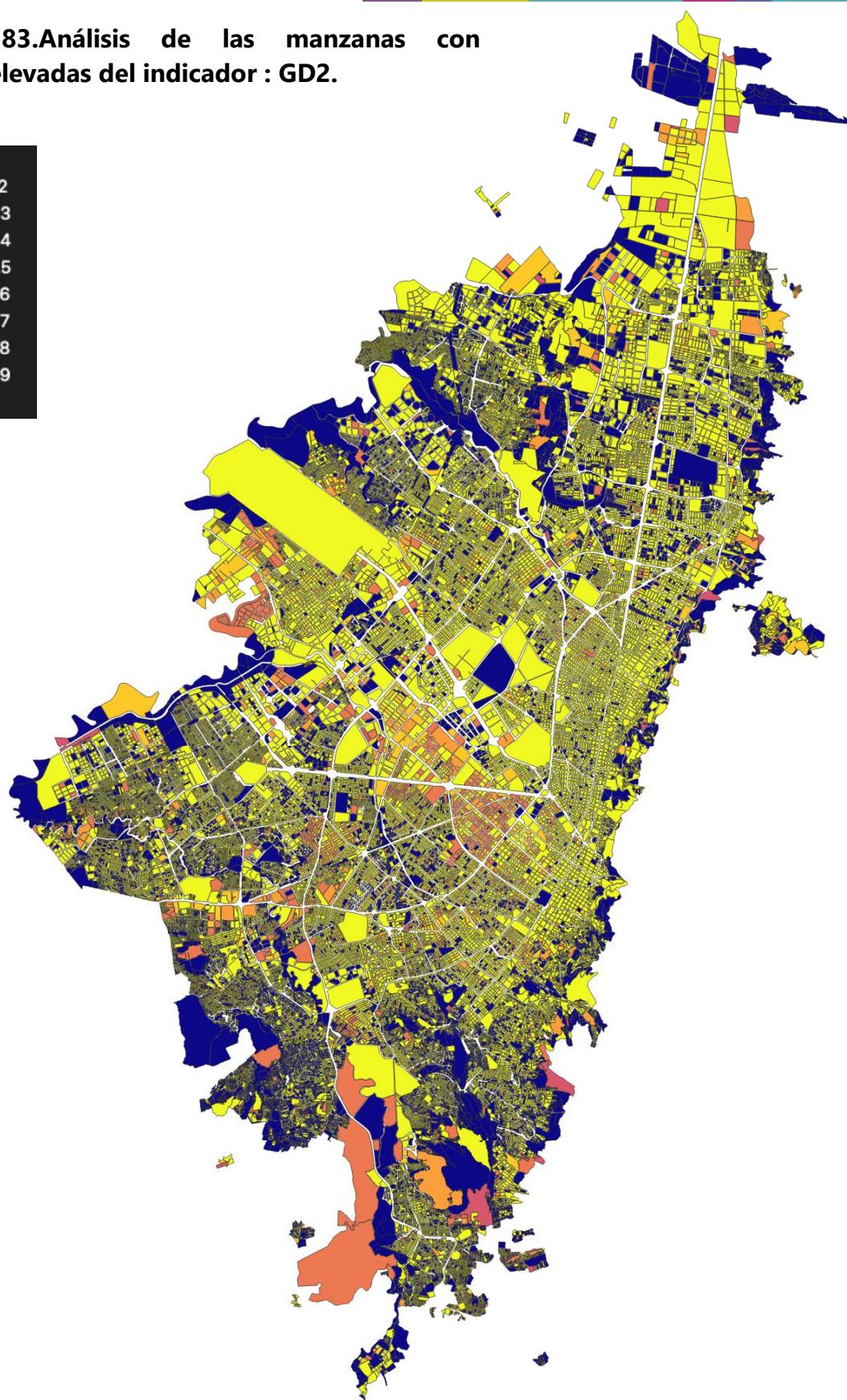
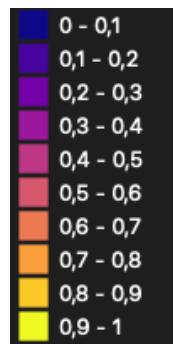


Ilustración 82. Mapas comparativos de significancia estadística con un 95% de confiabilidad ColdSpots

Para las manzanas en "Azul" (Método LISA) representadas anteriormente hacen referencias a que contemplando todas las permutaciones posibles y luego de un proceso aleatorio inminentemente estas zonas tengan este patrón de agrupación, en el que la perspectiva, se referencia a que conforme a las condiciones de aleatoriedad de la distribución de los datos del indicador cuando se muestran estos puntos "ColdSpots", estos hacen referencias a que las observaciones con perfiles similares en términos estadísticos están separadas gradualmente y además que no representan un peso estadísticamente significativo que permita agruparse entre ellos, pero que en este caso para la localidad de "Ciudad Bolívar", se muestra una alta concentración de manzanas que en sus agrupaciones reflejan valores de contigüidad bajos para la medición de las manzanas discriminadas en la anterior visualización y su vecindad refiere de valor bajos del indicador, es decir con una notable baja garantía de desarrollo por sector económico, es decir los clústeres en que los valores del indicador son bajos, es posible afirmar que en el 95% de los casos las cantidades unidades improductivas son significativamente mayores a las cantidades de unidades productivas en la localidad mencionada.

Ilustración 83. Análisis de las manzanas con mediciones elevadas del indicador : GD2.

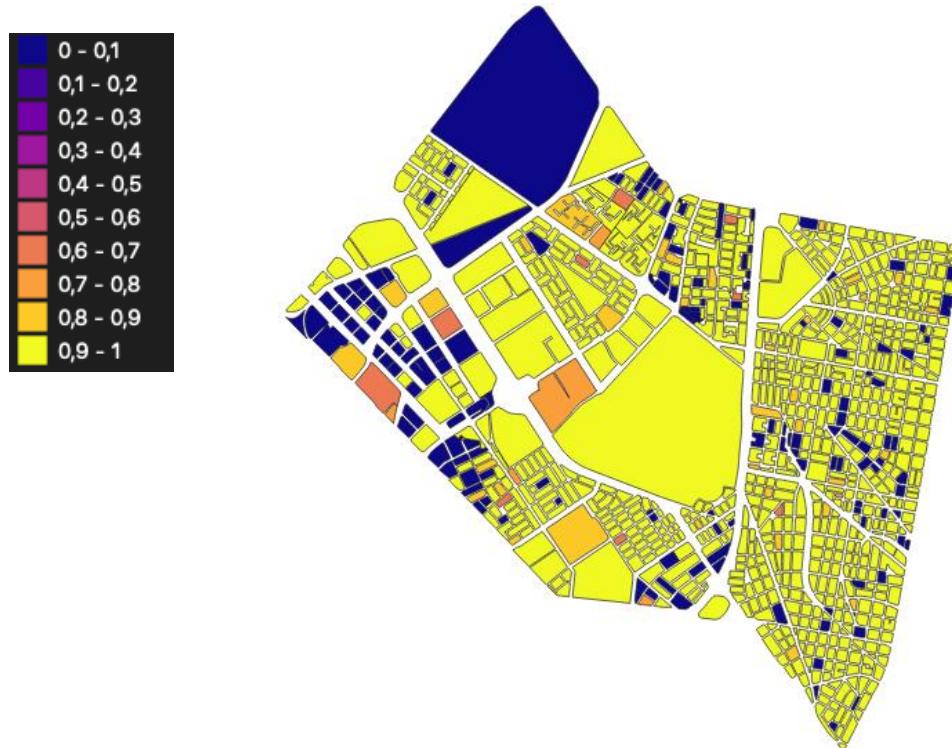


En la clasificación del indicador en la ciudad de Bogotá DC por manzana, es de notarse la distribución de las manzanas conforme a sus colores dados por la clasificación una alta presencia de manzanas con colores amarillo (valores entre 0.9 y 1).

Análisis de la localidad de Teusaquillo con mediciones del indicador :

Un vez revisado de forma general las comparaciones entre los análisis Geoestadístico en la ciudad de Bogotá DC en cada una de las localidades, se procede a analizar las incidencias del valor de indicador en cada manzana de la localidad de "Teusaquillo" en la ciudad de Bogotá, procesando así clasificaciones para diferentes escalas del indicador medido:

Ilustración 84. Mapa localidad Teusaquillo con distribución geoespacial del indicador GD2.

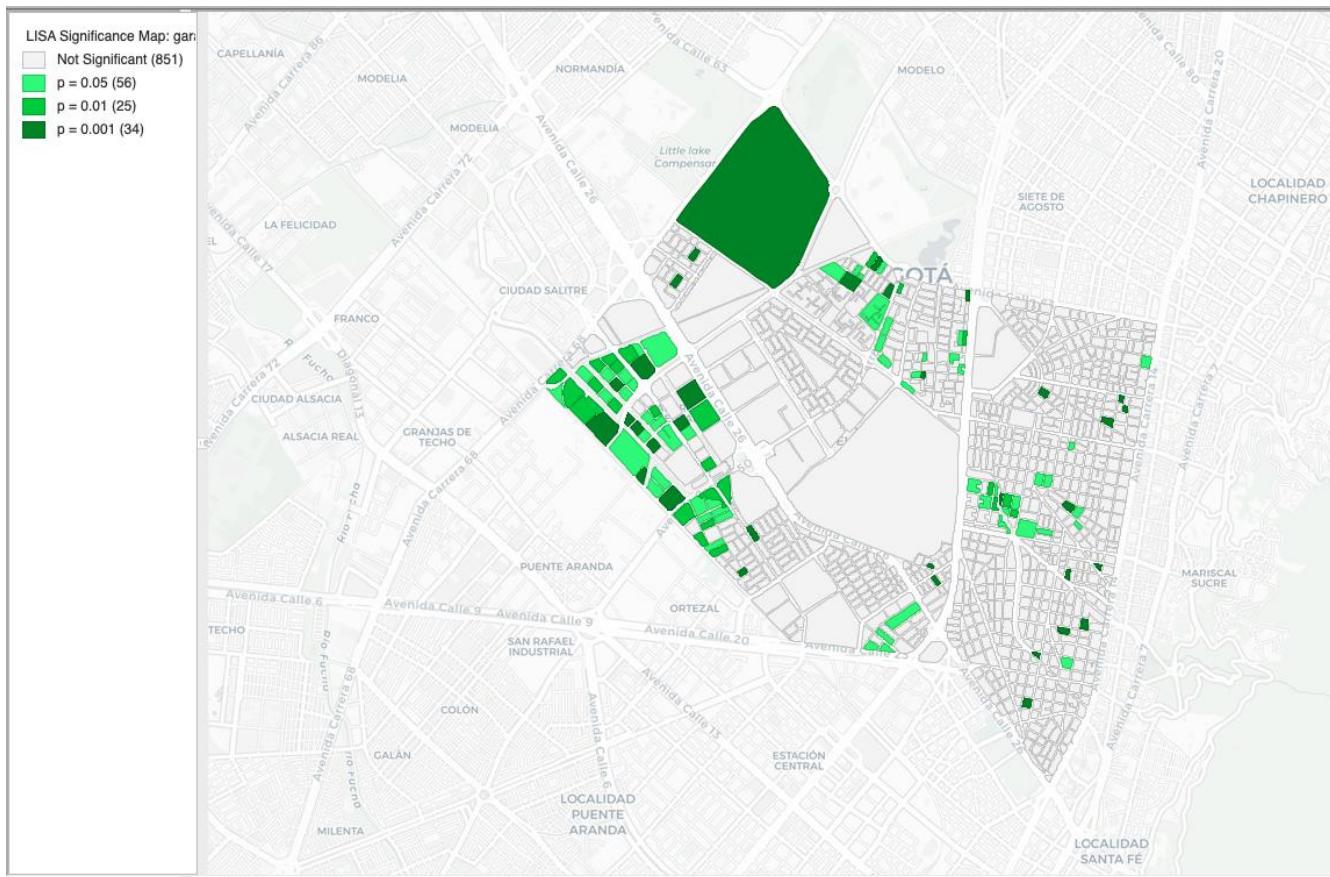


Fuente: Data set Ciudad Bogotá, MGN-NIM2018. Elaborado desde QGIS, Observatorio Inmobiliario Nacional DANE. Para una selección en término medio de la localidad, es de notarse la distribución de las manzanas conforme a sus colores dados por la clasificación, notándose alta presencia de manzanas con colores "Amarillos", estas ubicadas en el rango de [0.9,1] es decir que la proporción de unidades productivas se infiere una perfecta participación del sector económico terciario (de servicios).

Análisis prospectivo geoespacial de la localidad Teusaquillo:

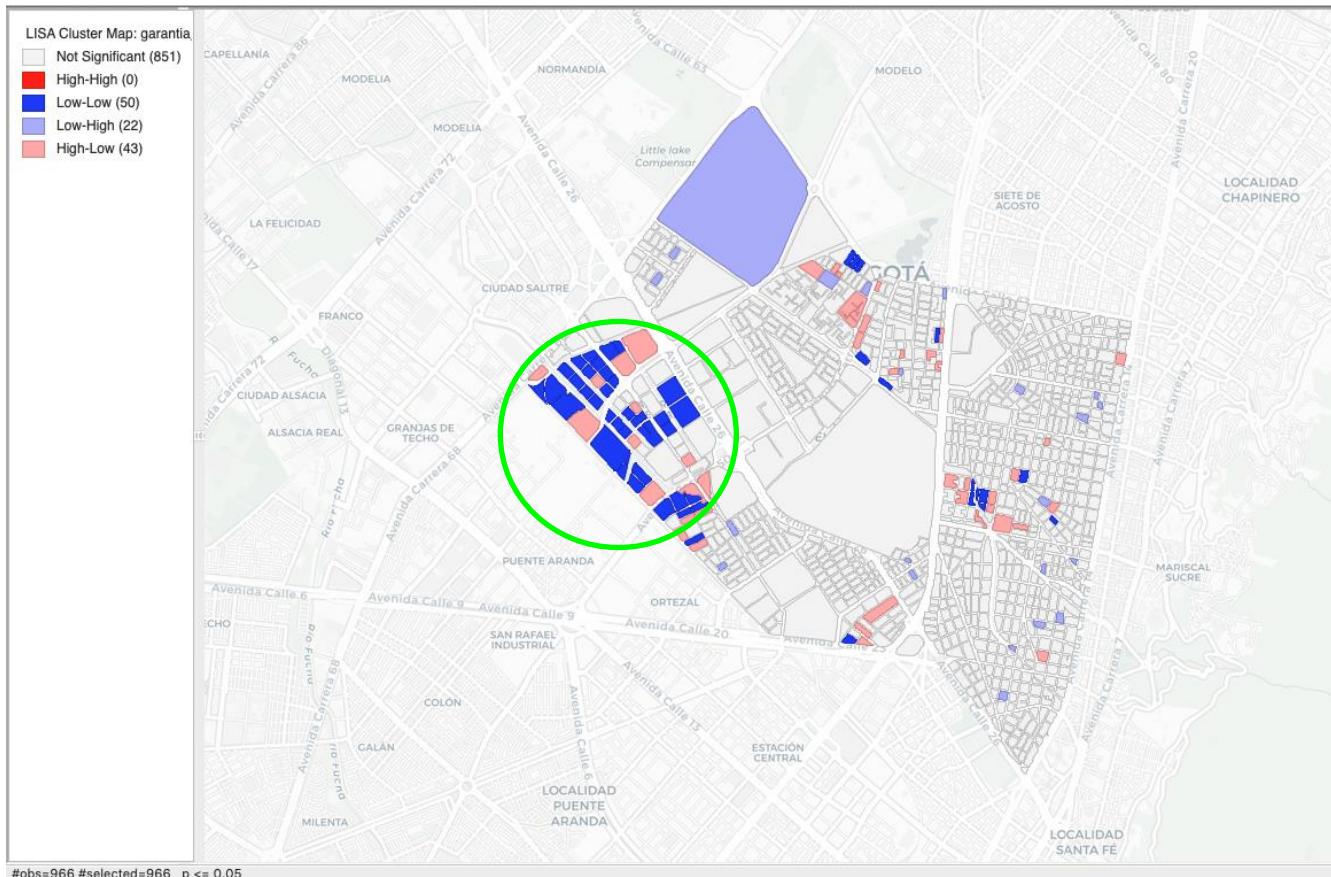
En la revisión geo-estadística del indicador en cuestión, para la localidad de Teusaquillo conforme a las características de distribución de los datos, es posible apreciar el desarrollo de los niveles de significancia estadística para cada manzana dentro de los perímetros de la localidad. Gracias a este método del Índice de Morán se generan los siguientes de mapas de significancia estadística.

Ilustración 85. Mapa de significancia método LISA, 95% confiabilidad para el indicador GD2.



Las manzanas que muestra el mapa, son aquellas que sus valores no se deben a instancias aleatoria, sino que existe una fuerte influencia de su ubicación espacial para que estas tengan valores ya sean elevados del indicador o bajos del indicador de garantía de desarrollo por sector económico.

Ilustración 86. Mapa de clúster por el método LISA con peso de contigüidad KNN para el indicador GD2.



En el 95% de los casos es probable encontrar que las manzanas en los barrios de "Ciudad Salitre Oriental" y "Quinta Paredes" (identificados en círculo verde 0) que se agrupen en algún tipo de clúster con valores con bajos niveles de garantía de desarrollo productivo, por tener mayor niveles que no representan dentro del concepto de la garantía de desarrollo productivo alguna manifestación de los sectores de la economía, esto se debe a factores de la naturaleza de los predios, este grupo de manzanas en color azul representan los "ColdSpots". En la localidad de Teusaquillo no existen zonas en las que se puedan considerar agrupaciones fiables de manzanas con valores significativos del indicador.

Trabajos citados

- Rodrigo, J. A. (2021). *Analisis de normalidad*. Obtenido de ciencia de datos :
https://www.cienciadedatos.net/documentos/8_analisis_normalidad
- Mitjana, L. R. (2021). *Prueba Chi-cuadrado Usos estadísticos* . Obtenido de Psicología y mente:
<https://psicologiyamente.com/misclanea/prueba-chi-cuadrado>
- Rodrigo, J. A. (2021). *Análisis de la Homogenidad de varianza (Homocesdasticidad)*. Obtenido de ciencia de datos:
https://www.cienciadedatos.net/documentos/9_homogeneidad_de_varianza_homocedasticidad.html#Conclusión
- ArcGIS . (2021). *Funcionamiento autocorrelación espacial Indice de Morán*. Obtenido de ArcGIS:
<https://desktop.arcgis.com/es/arcmap/latest/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>
- Vicens, L. (6 de Mayo de 2021). *Introducción a GEODA: Analisis de autocorrelación espacial univariada*. Obtenido de <https://www.youtube.com/watch?v=AdAAtcVKchA&t=598s>
- GEASIG. (2021). *Análisis de patrones espaciales*. Obtenido de GEASIG:
<https://www.geasig.com/analisis-de-patrones-espaciales-con-arcgis/>
- Chavez, P. R. (4 de Julio de 2021). *Visualizando la información de las tablas cruzadas*. Obtenido de Aprende Ciencia de Datos: <https://www.youtube.com/watch?v=rdvpRAhYTVY>
- Chavez, R. (21 de Mayo de 2020). *Analís exploratorio con ACP en RStudio*. Obtenido de Aprende Ciencia de Datos: <https://www.youtube.com/watch?v=teqJOyyVSpA>
- Minitab Blog Editor. (18 de Mayo de 2016). *Understanding Analysis of Variance (ANOVA) and the F-test*. Obtenido de MiniTab Blog: <https://blog.minitab.com/en/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test>
- UCLA. (s.f.). *WHAT DOES CRONBACH'S ALPHA MEAN? | SPSS FAQ*. Obtenido de <https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/>
- SAGE JOURNALS. (6 de Noviembre de 2020). *Advances in Methods and Practices in Psychological Science*. Obtenido de
<https://journals.sagepub.com/doi/full/10.1177/2515245920951747>