

Consumo de agua - CDMX

Equipo 7



GOBIERNO DE LA
CIUDAD DE MÉXICO

Datos Abiertos Ciudad de México

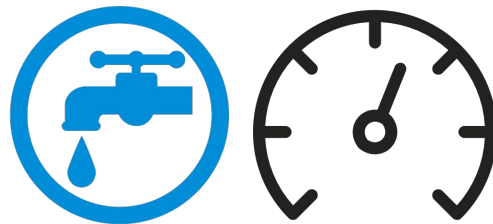
¿Cómo se mide el consumo de agua en México?

Información bimestral

Suministro de agua a nivel manzana

Facturación por servicio de consumo medido y promedio

Consumo medido de un medidor instalado/aprobado por el sistema de Aguas



Fuente: [portal de datos abiertos de la CDMX](#)



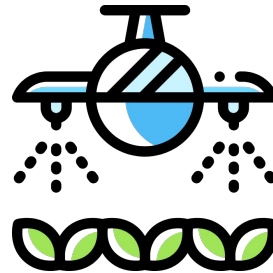
GOBIERNO DE LA
CIUDAD DE MÉXICO

Uso del consumo de agua en México

Doméstico: Inmuebles de uso Habitacional

No doméstico: Inmuebles de uso no Habitacional

Mixto: Inmuebles de uso Habitacional y no Habitacional simultáneamente



Fuente: Código Fiscal de la Ciudad de México ejercicio 2019



GOBIERNO DE LA
CIUDAD DE MÉXICO



Data profiling → Datos crudos

Estadísticas	Resultado
Total de variables	17
Conteo de observaciones	71,102
Total de celdas	1,208,734
Cantidad de variables numericas	11
Cantidad de variables de fecha	0
Cantidad de variables categóricas	0
Cantidad de variables de texto	6
Valores faltantes	26,318
Porcentaje de valores faltantes	2.2%
Renglones duplicados	0
Porcentaje de valores duplicados	0.0%

Variables

```
Geo Point
Geo Shape
consumo_total_mixto
anio
nomgeo
consumo_prom_dom
consumo_total_dom
alcaldia
colonia
consumo_prom_mixto
consumo_total
consumo_prom
consumo_prom_no_dom
bimestre
consumo_total_no_dom
gid
indice_des
```





Data profiling → Revisión

número de elementos únicos	
Geo Point	22,930
Geo Shape	22,922
consumo_total_mixto	24,339
anio	1
nomgeo	17
consumo_prom_dom	52,060
consumo_total_dom	47,051
alcaldia	16
colonia	1,340
consumo_prom_mixto	31,911
consumo_total	56,015
consumo_prom	62,214
consumo_prom_no_dom	37,440
bimestre	3
consumo_total_no_dom	27,336
gid	71,102
indice_des	4

Datos únicamente de 2019

Hay 3 bimestres: 1, 2 y 3

Se detectó un error en nomgeo → Talpan

Nomgeo y alcaldia es la misma informacion

Hay 4 indices de desarrollo: Popular, bajo, medio y alto

1340 colonias



GOBIERNO DE LA
CIUDAD DE MÉXICO

Data profiling → NA's

	Missing Values	% del Total
consumo_total_mixto	8,327	11.7
consumo_prom_mixto	8,327	11.7
consumo_prom_dom	4,820	6.8
consumo_total_dom	4,820	6.8
Geo Shape	24	0.0

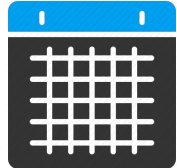
8908 renglones con al menos un valor nulo

12.53% del total de renglones

Máximo de registros nulos por renglón: 5

	renglon	valores_nulos
54555	Nan in row 54555	5
13306	Nan in row 13306	4
24048	Nan in row 24048	4
24061	Nan in row 24061	4
52584	Nan in row 52584	4
24052	Nan in row 24052	4
55696	Nan in row 55696	4
24050	Nan in row 24050	4
24049	Nan in row 24049	4
52583	Nan in row 52583	4





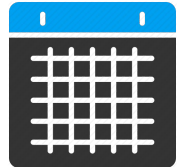
Data profiling → Variables numéricas

	metrica	consumo_total	consumo_total_mixto	consumo_total_dom	consumo_total_no_dom
0	tipo	float64	float64	float64	float64
1	numero de observaciones	71102	62775	66282	71102
2	media	1695.85	174.36	1186.26	436.06
3	desviacion estándar	3555.7	312.66	2771.04	2126.15
4	cuartil 25%	340.95	0	161.64	10.98
5	cuartil 50%	896.18	79.94	604.18	54.06
6	cuartil 75%	1808.9	233.32	1261.45	230.43
7	mínimo	0	0	0	0
8	máximo	119727	23404.4	95060.7	119727
9	numero de observaciones únicas	56015	24339	47051	27336

Distribuciones sesgadas a la izquierda con alta dispersión



GOBIERNO DE LA
CIUDAD DE MÉXICO



Data profiling → Variables numéricas

	metrica	consumo_prom	consumo_prom_mixto	consumo_prom_dom	consumo_prom_no_dom
0	tipo	float64	float64	float64	float64
1	numero de observaciones	71102	62775	66282	71102
2	media	111.22	50.64	29.13	126.76
3	desviacion estándar	1069.95	130.41	64.57	1095.82
4	cuartil 25%	23.01	0	18.69	6.28
5	cuartil 50%	31.69	33.45	26.41	19.28
6	cuartil 75%	45.48	61.22	36.25	54.19
7	minimo	0	0	0	0
8	maximo	89691.8	11702.2	7796.41	89691.8
9	numero de observaciones unicas	62214	31911	52060	37440

Distribuciones sesgadas a la izquierda con alta dispersión



GOBIERNO DE LA
CIUDAD DE MÉXICO



Data profiling → Variables categóricas

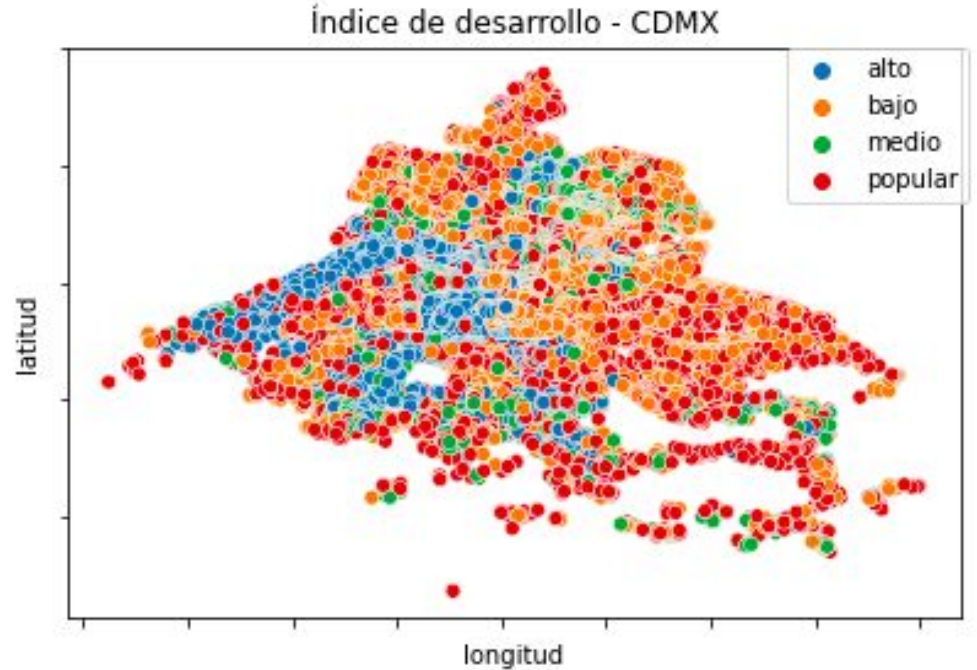
	metrica	nomgeo	alcaldia	colonia	bimestre	gid	indice_des
0	tipo	category	category	category	category	category	category
1	numero de categorias	16	16	1340	3	71102	4
2	numero de observaciones	71102	71102	71102	71102	71102	71102
3	observaciones nulas	0	0	0	0	0	0
4	% observaciones nulas	0	0	0	0	0	0
5	valores unicos	[gustavo a. madero, azcapotzalco, coyoacan, ti...	[gustavo a. madero, azcapotzalco, coyoacan, ti...	[7 de noviembre, nueva tenochtitlan, prohogar,...	[3, 1, 2]	[57250, 57253, 57255, 57267, 57330, 57273, 572...	[alto, medio, popular, bajo]
6	moda1/veces/porcentaje	[iztapalapa, 10515, 14.79%]	[iztapalapa, 10515, 14.79%]	[centro, 1139, 1.6%]	[2, 23942, 33.67%]	[71102, 1, 0.0%]	[bajo, 29248, 41.14%]
7	moda2/veces/porcentaje	[gustavo a. madero, 10058, 14.15%]	[gustavo a. madero, 10058, 14.15%]	[agricola oriental, 837, 1.18%]	[3, 23822, 33.5%]	[23703, 1, 0.0%]	[popular, 16539, 23.26%]
8	moda3/veces/porcentaje	[cuauhtemoc, 7313, 10.29%]	[cuauhtemoc, 7313, 10.29%]	[roma norte, 602, 0.85%]	[1, 23338, 32.82%]	[23697, 1, 0.0%]	[alto, 15516, 21.82%]





Data profiling → Índice de desarrollo

	indice_des	porcentaje
bajo	29,248	41.14%
popular	16,539	23.26%
alto	15,516	21.82%
medio	9,799	13.78%





Data Profiling → Datos limpios

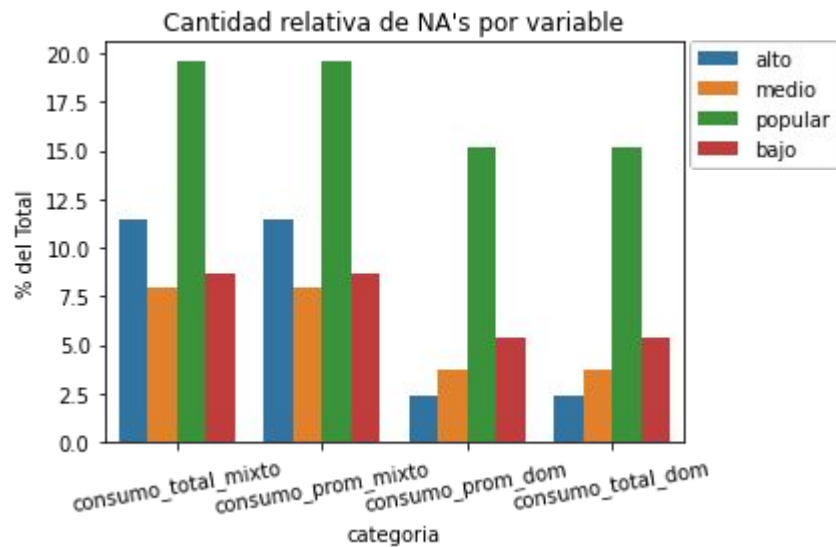
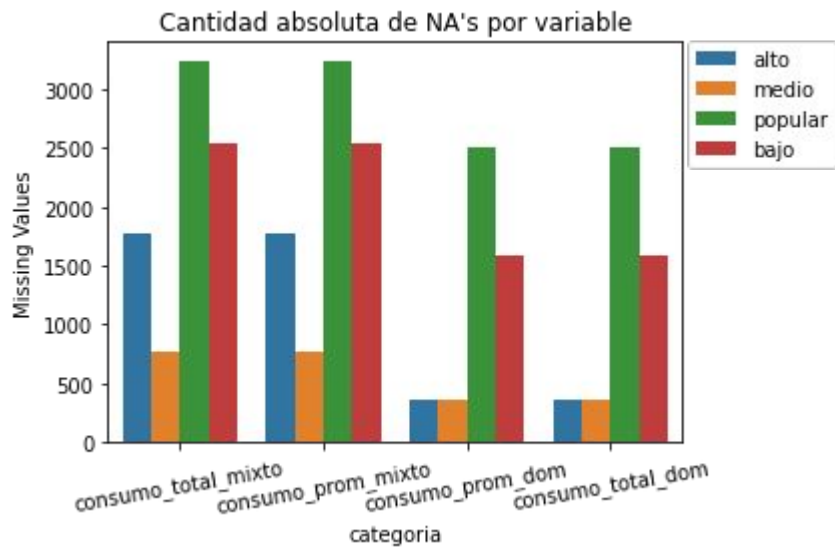
Estadísticas	Resultado
Total de variables	17
Conteo de observaciones	71,102
Total de celdas	1,208,734
Cantidad de variables numericas	11
Cantidad de variables de fecha	0
Cantidad de variables categóricas	6
Cantidad de variables de texto	0
Valores faltantes	26,294
Porcentaje de valores faltantes	2.2%
Renglones duplicados	0
Porcentaje de valores duplicados	0.0%

Ajuste y limpieza de variables:

- Corrección de datos erróneos
- Identificación de variables redundantes
- Entendimiento de dispersión
- Valorar duplicidad en información

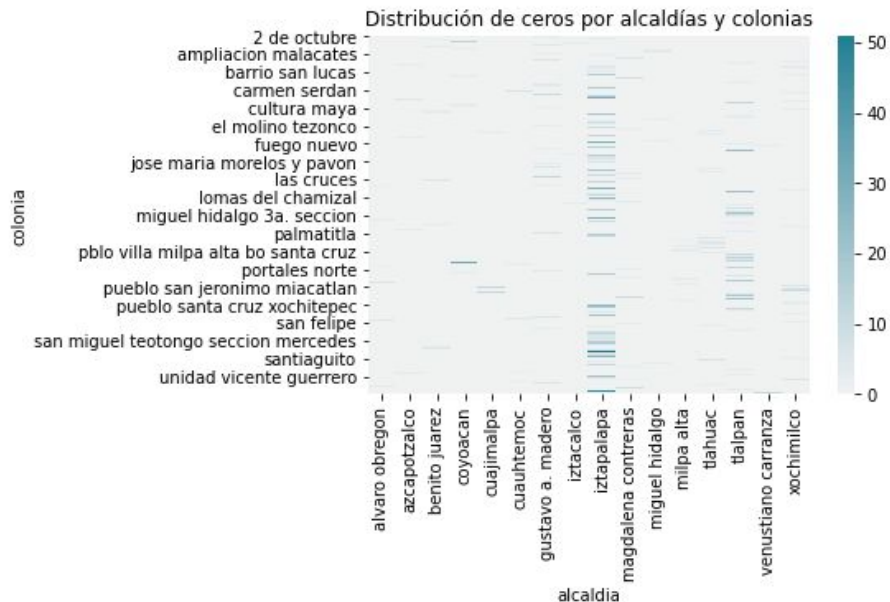


¿Cómo están los NA's?



¿Cómo están distribuidos los ceros?

- Las observaciones sin información representan 3.4% de los datos
- Las alcaldías con mayor cantidad de ceros son Iztapalapa, seguida de Tlalpan



¿Cómo están distribuidos los ceros?

Datos originales

alcaldia	colonia
iztapalapa	193
alvaro obregon	188
gustavo a. madero	167
tlalpan	130
coyoacan	96
xochimilco	90
azcapotzalco	88
miguel hidalgo	86
tlahuac	70
venustiano carranza	67
benito juarez	53
cuajimalpa	39
magdalena contreras	38
iztacalco	38
cuauhtemoc	35
milpa alta	33

Datos cero

alcaldia	colonia
iztapalapa	90
tlalpan	56
gustavo a. madero	32
xochimilco	26
tlahuac	22
magdalena contreras	20
alvaro obregon	18
coyoacan	17
milpa alta	11
benito juarez	11
miguel hidalgo	9
cuauhtemoc	8
azcapotzalco	8
venustiano carranza	7
cuajimalpa	6
iztacalco	5

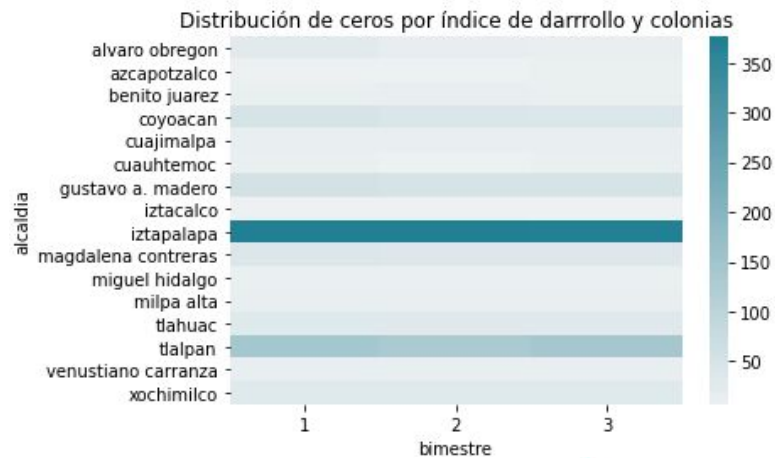
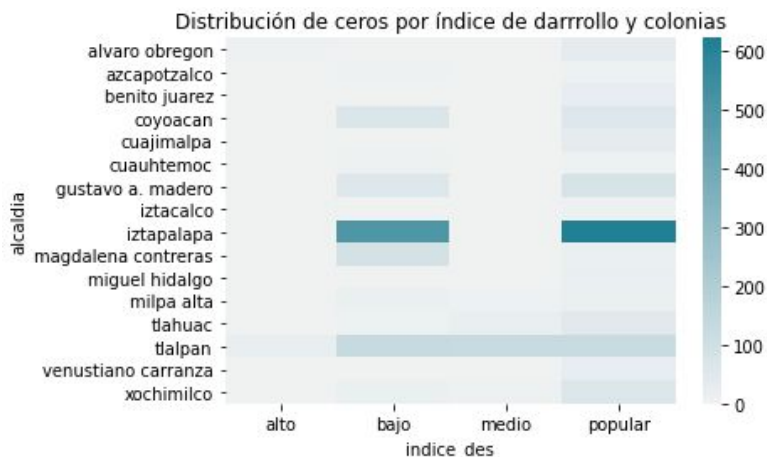
- Los datos no son afectados por este suceso de manera uniforme, pues se esperaría que se mantuviera el orden de las alcaldías en relación a la cantidad de colonias que poseen.
- Hay colonias que son más afectadas por otras habrá que aclarar a qué se debe este fenómeno.





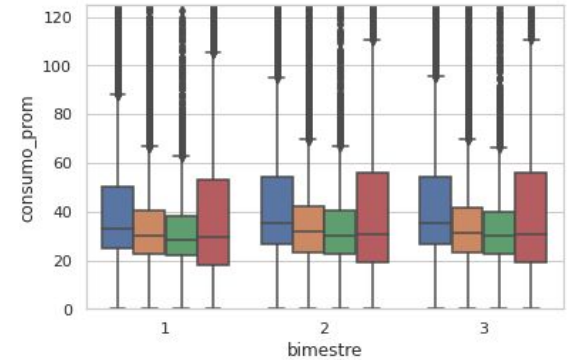
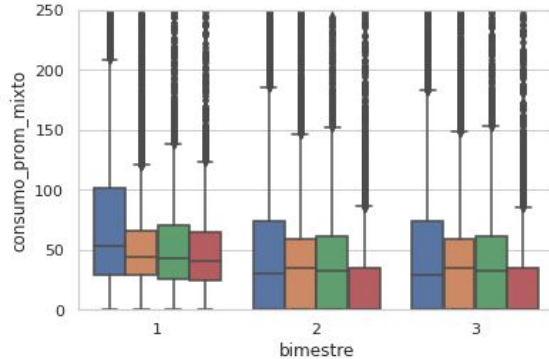
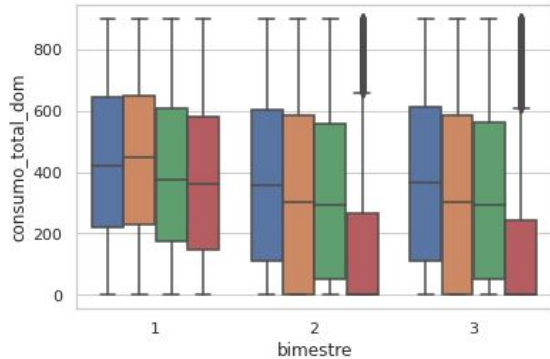
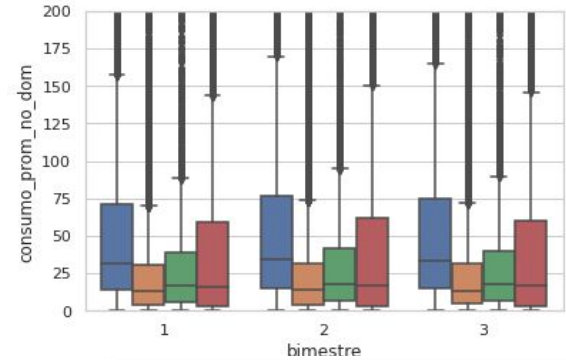
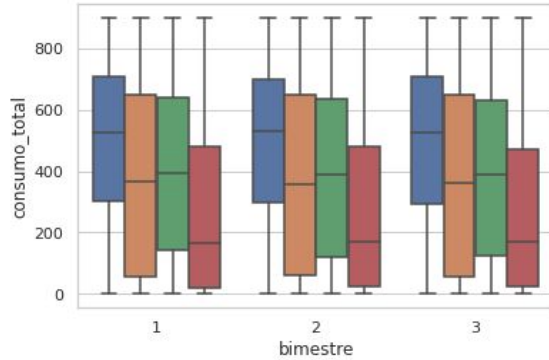
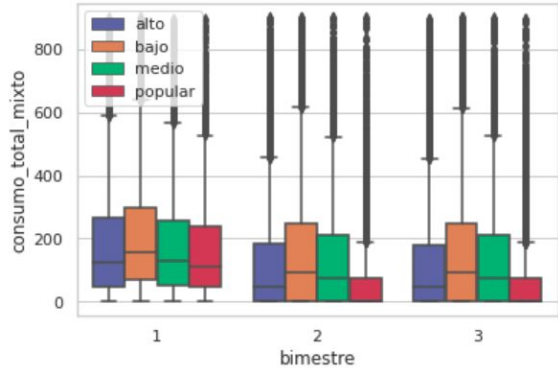
¿Cómo están distribuidos los ceros?

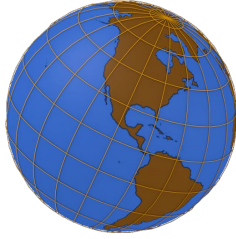
- A lo largo del tiempo, las distintas alcaldías han tenido prácticamente la misma cantidad de 0's en el total de las observaciones
- Los principales índices de desarrollo con ceros en las observaciones son popular, seguido del bajo





¿Existe algún cambio entre bimestres?





¿Tomar alcaldías o coordenadas?

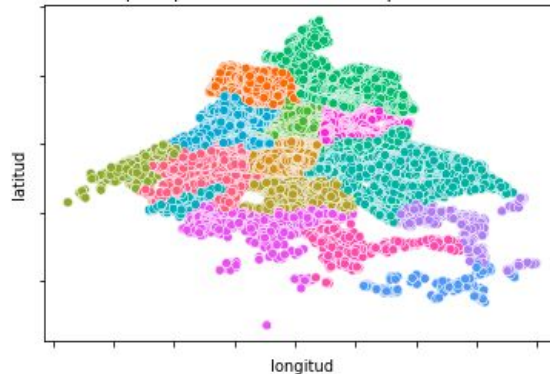
Alcaldías

- Asume que el comportamiento de las observaciones dentro de la alcaldía es similar
- Ayuda a simplificar modelos, teniendo únicamente 16 puntos geospaciales (alcaldías)
- La variable alcaldía es categórica, que será transformada a dummie al momento de utilizarla en los futuros modelos, lo cual hará que la dimensión de nuestro dataset incremente automáticamente en 15

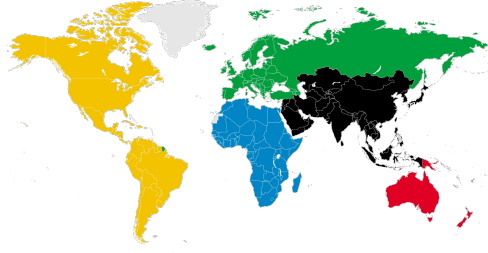
Coordenadas

- Permite encontrar diferencias entre alcaldías
- Se tendrán tantos puntos geospaciales como observaciones (asumiendo que cada observación tiene distinta ubicación)
- Al tratar con puntos geospaciales, únicamente se utilizarán las variables latitud y longitud en el modelo, lo cual puede simplificar la estructura del dataset

Mapa representativo de CDMX por alcaldías



¿Tomar coordenadas o alcaldías?



Con base en los comentarios anteriores, se llegará a una conclusión con el cliente haciendo hincapié en los siguientes puntos:

- Alcaldías.- su principal ventaja es que se enfoca en 16 puntos únicos (alcaldías)
- Coordenadas.- logra obtener mayor profundidad, ayudando a encontrar diferencias dentro de las alcaldías

Se probará el/los modelo(s) con alcaldías y coordenadas y se elegirá la opción que ayude más a cumplir con el/los objetivo(s)

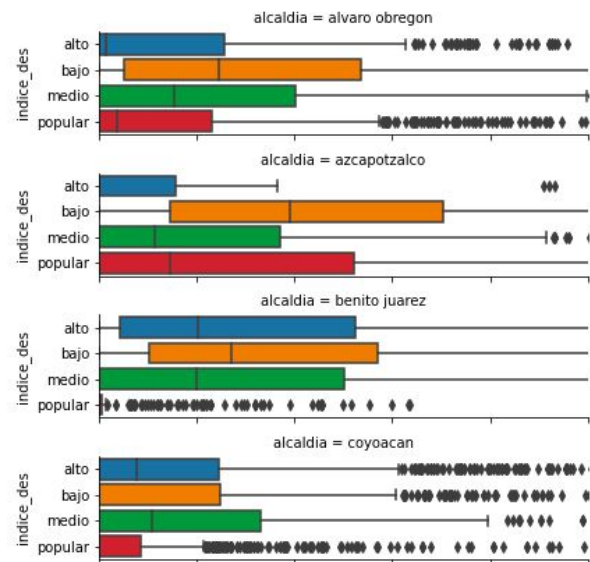


GOBIERNO DE LA
CIUDAD DE MÉXICO

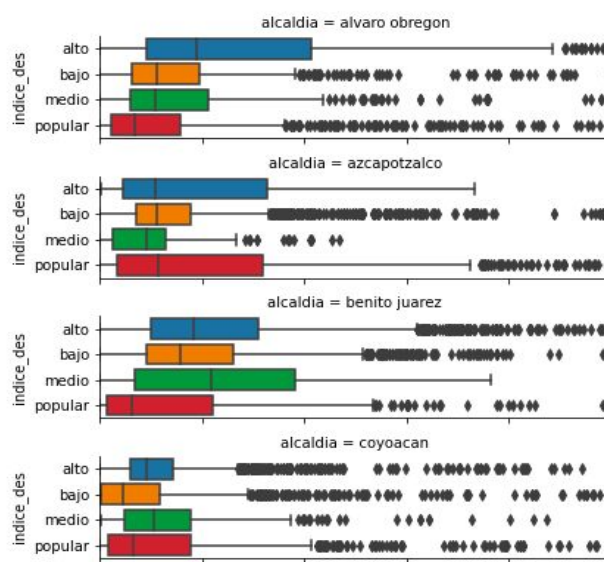


¿Tomar coordenadas o alcaldías?

Consumo total mixto por índice de desarrollo



Consumo total por índice de desarrollo

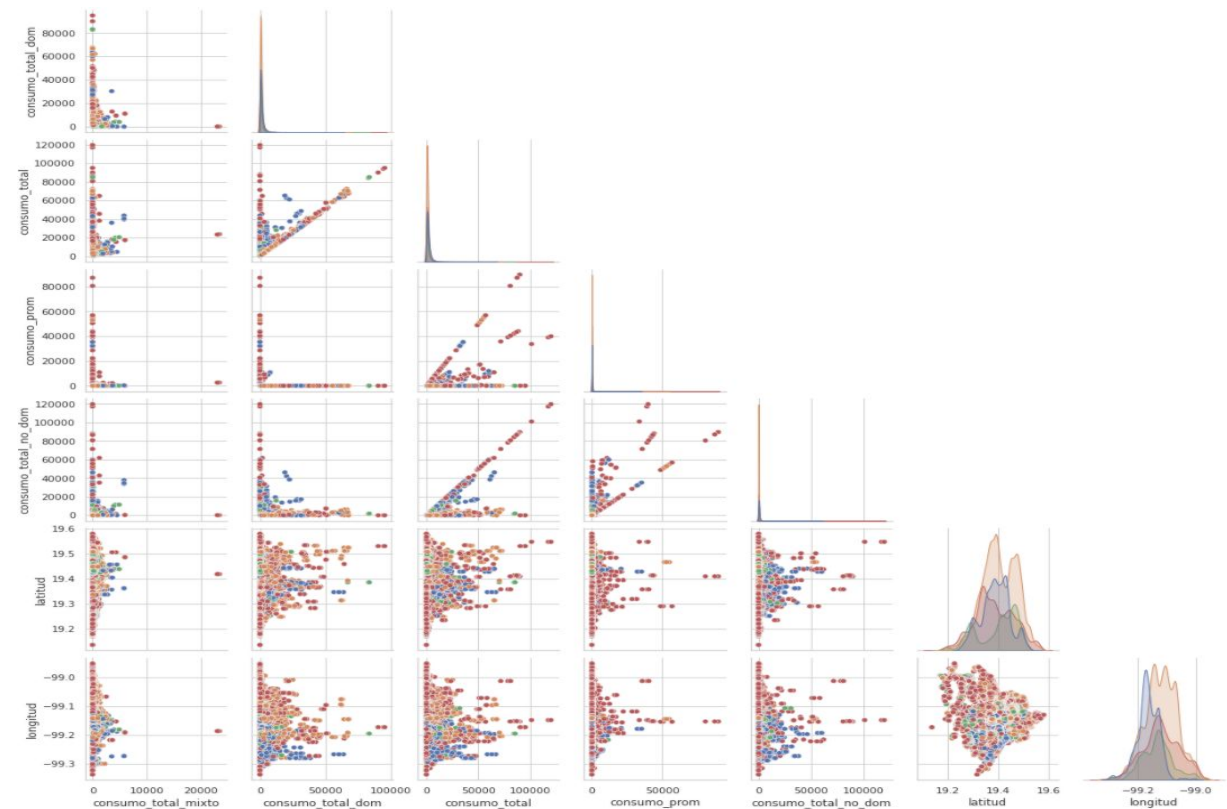


Ejemplo de cómo a nivel alcaldía las variables de Consumo total y Consumo total mixto pueden identificar diferentes patrones entre cada una de las alcaldías.



GOBIERNO DE LA
CIUDAD DE MÉXICO

Variables seleccionadas



- Las variables geoespaciales son las que visualmente ayudan más a la distinción entre índices
- Las combinaciones mostradas ayudan relativamente a lograr el objetivo



Variables seleccionadas ¿por qué?

- ¿latitud?
- ¿longitud?
- ¿alcaldia?

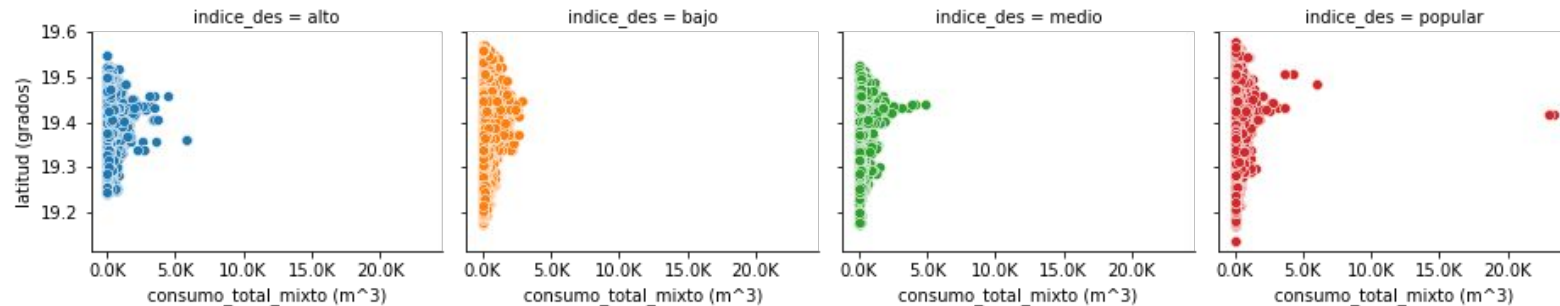
probar el modelo con alcaldia vs lat-long

-
- consumo_total_mixto
 - consumo_total_dom
 - consumo_total
 - consumo_total_no_dom
 - consumo_prom

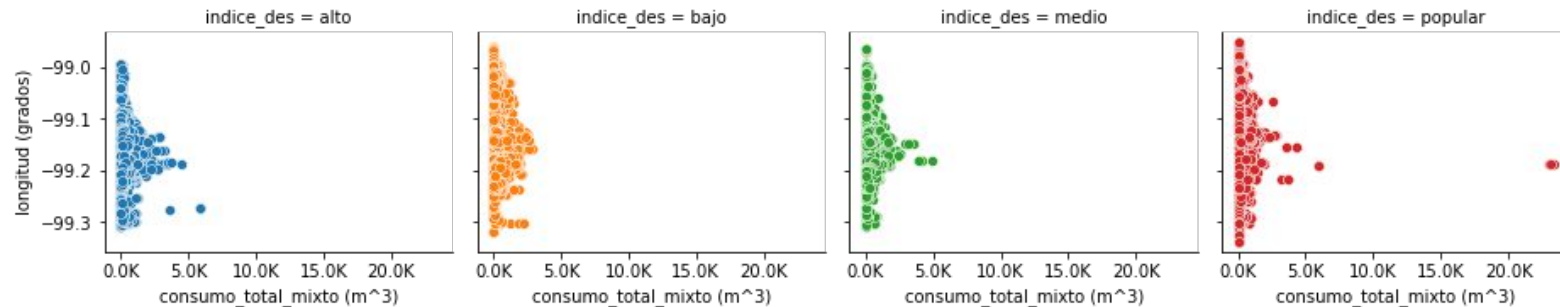


GOBIERNO DE LA
CIUDAD DE MÉXICO

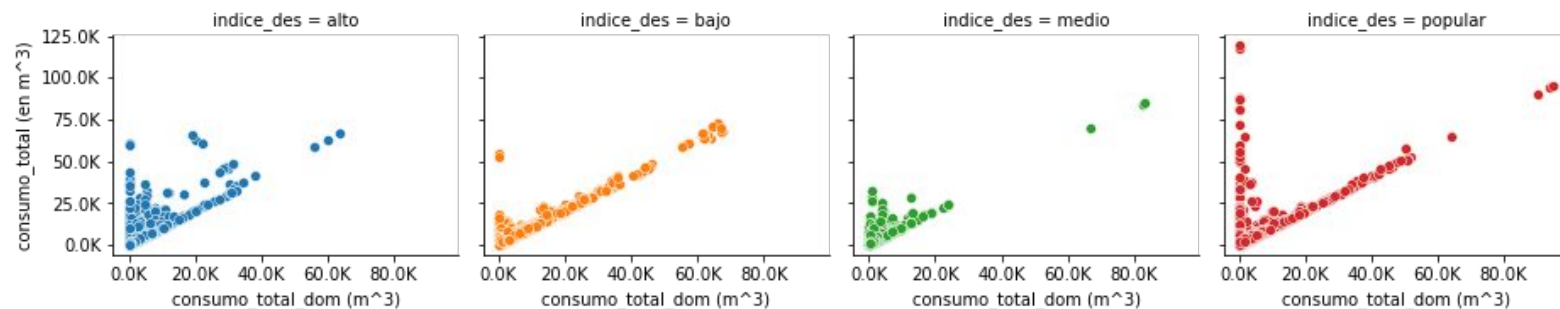
Consumo total mixto por latitud



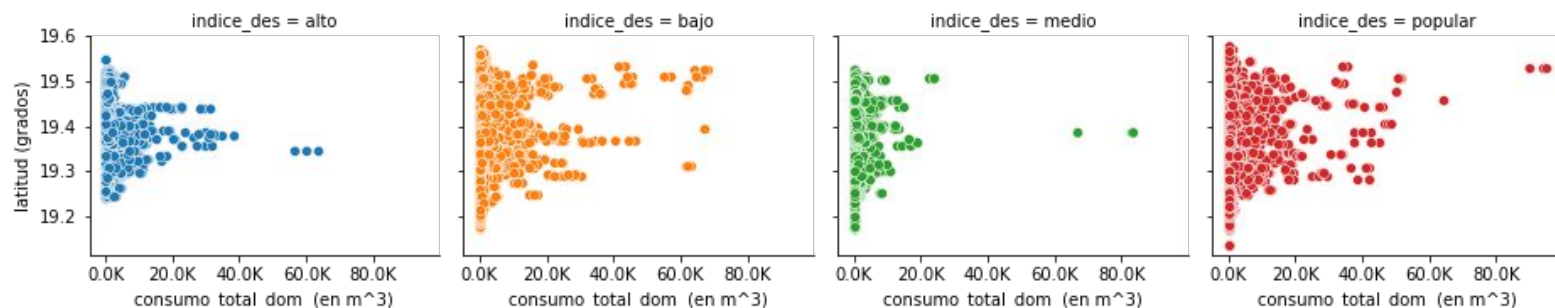
Consumo total mixto por longitud



Consumo total dom por consumo total

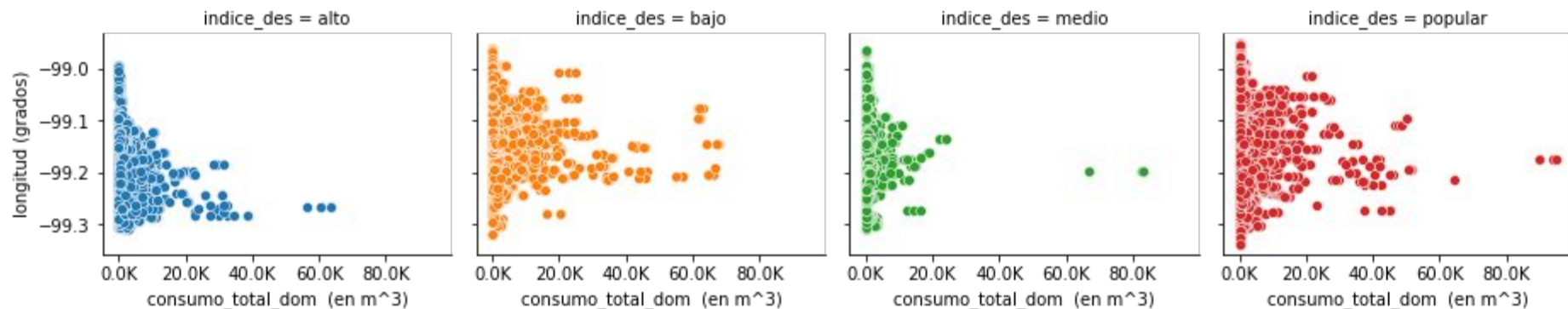


Consumo total dom por latitud

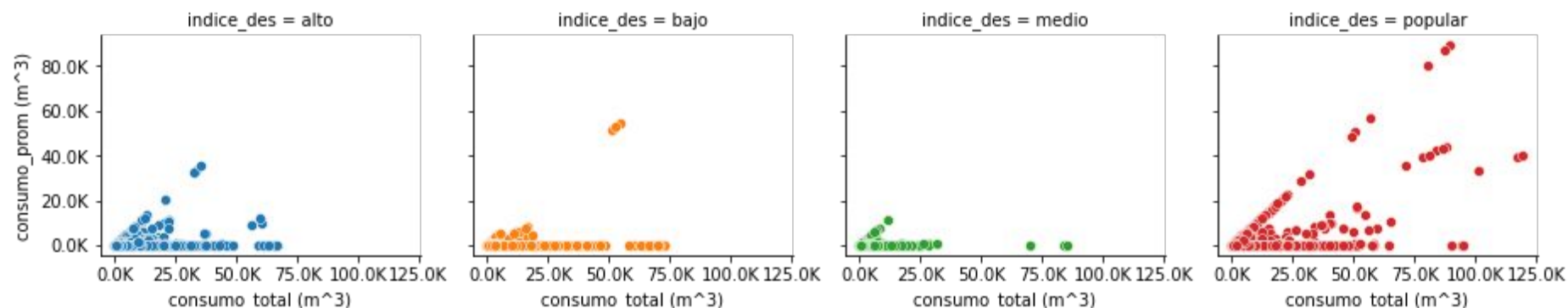


GOBIERNO DE LA
CIUDAD DE MÉXICO

Consumo total dom por longitud

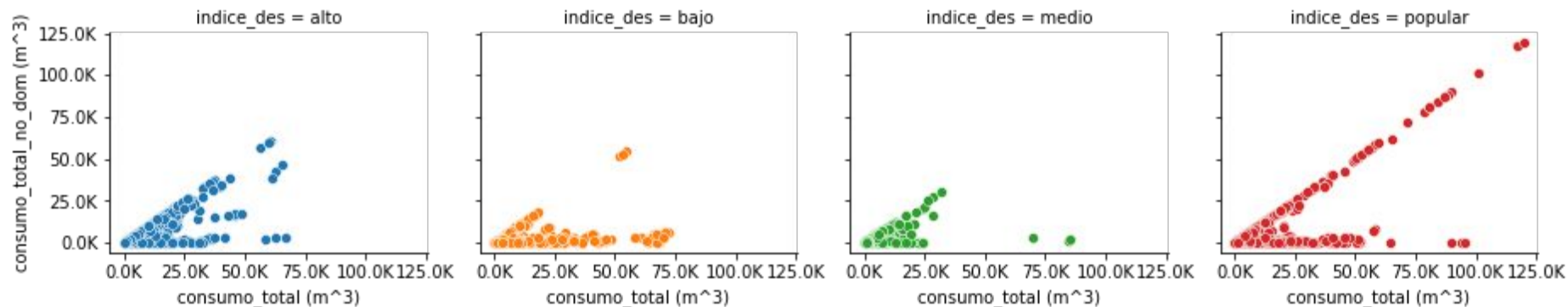


Consumo total por consumo promedio

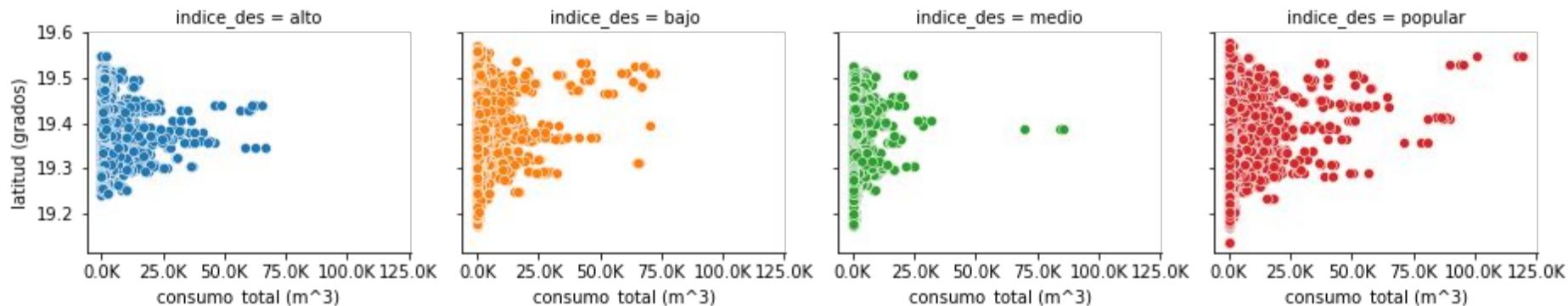


GOBIERNO DE LA
CIUDAD DE MÉXICO

Consumo total por consumo no dom

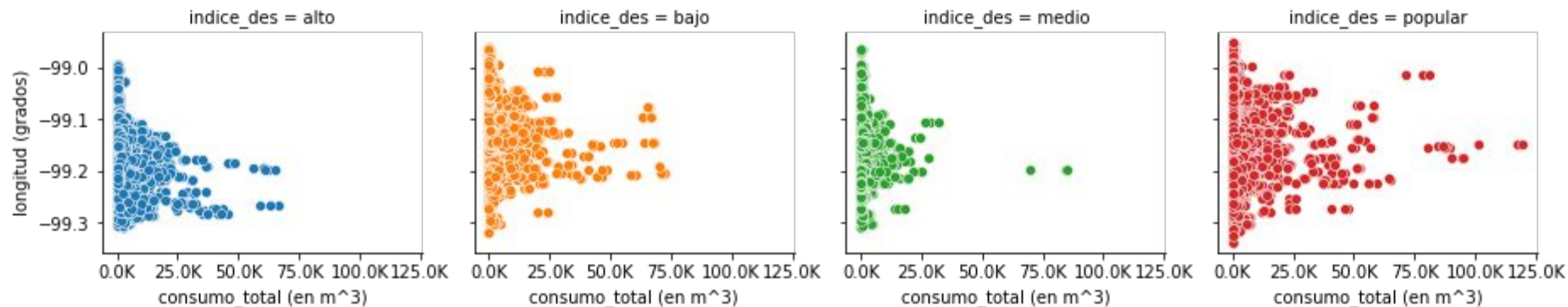


Consumo total por latitud

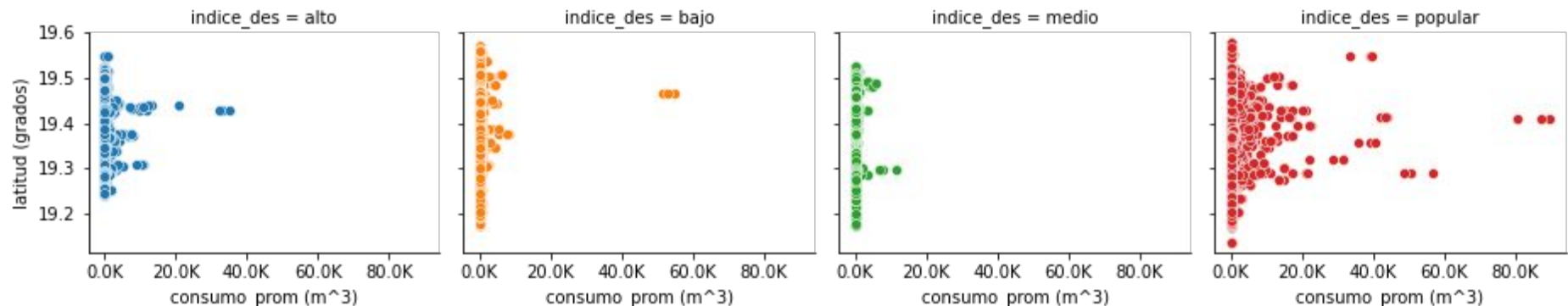


GOBIERNO DE LA
CIUDAD DE MÉXICO

Consumo total por longitud

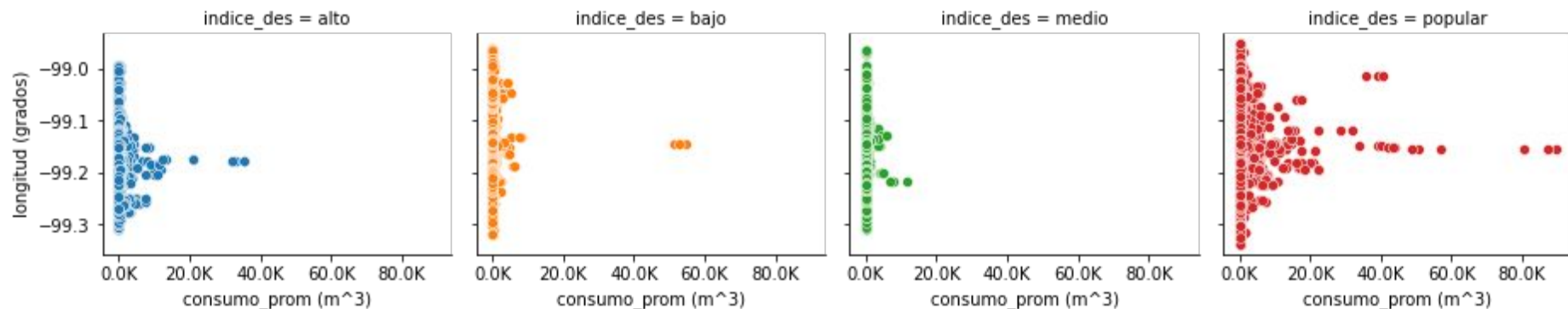


Consumo promedio por latitud

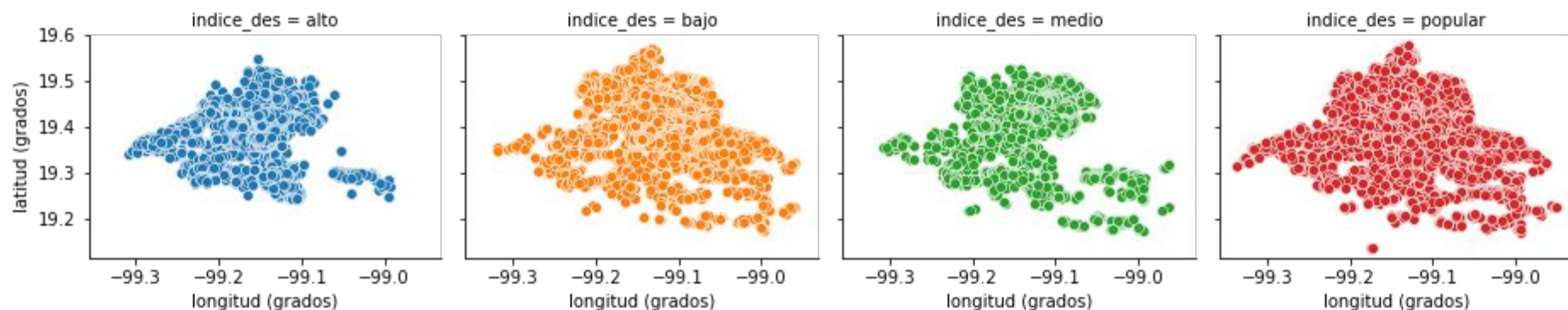


GOBIERNO DE LA
CIUDAD DE MÉXICO

Consumo promedio por longitud



longitud por latitud



GOBIERNO DE LA
CIUDAD DE MÉXICO

Variables seleccionadas ¿por qué?

- ¿latitud?
- ¿longitud?
- ¿alcaldia?

probar el modelo con alcaldia vs lat-long

-
- consumo_total_mixto
 - consumo_total_dom
 - consumo_total
 - consumo_total_no_dom
 - consumo_prom



GOBIERNO DE LA
CIUDAD DE MÉXICO

Consideraciones

Preguntas

- ¿Es diferente el consumo del segundo semestre del año?
- ¿Los consumos varían significativamente entre colonias cada año?
- ¿Qué tanto y cada cuánto corrigen los datos de las mediciones?
- No hay relación en valores omitidos: mixto, no dom y dom. ¿Es correcto?
- La cantidad de 0's se deben a la forma de medir o es un error en el proceso. ¿Es correcto?

Ética

- Representatividad de todos los inmuebles en la CDMX para que puedan participar equitativamente en programas de subsidios.



GOBIERNO DE LA
CIUDAD DE MÉXICO