

#	Column
0	business_id
1	name
2	address
3	state
4	latitude
5	longitude
6	stars
7	review_count
8	categories

8	Tip_Comida
9	DCP_FL
10	CRE_FL
11	DCP_FL_C
12	SEG_DCP_FL
13	Ind_DCP_FL_Prom
14	Ind_DCP_FL_Max
15	Prom_SEG_PEE_FL
16	Max_SEG_PEE_FL
17	PEE_FL_Cumple

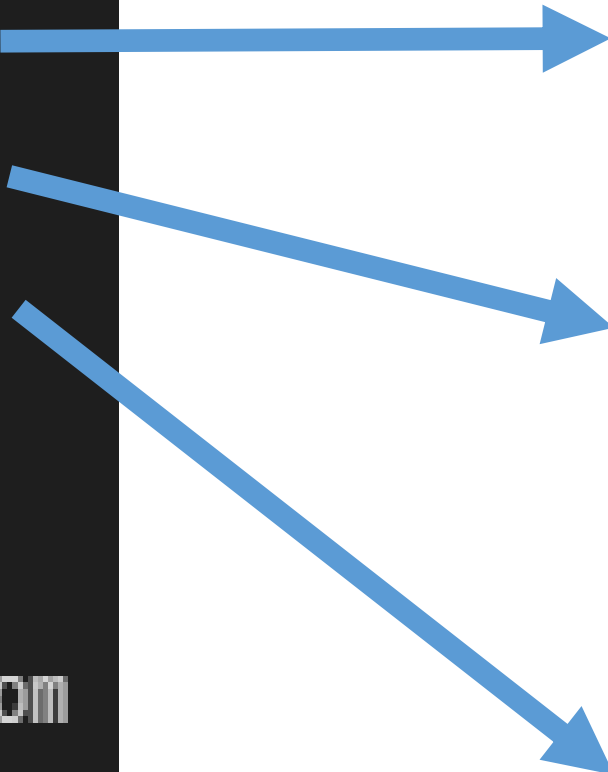
Atributos fuentes del archivo business.
De este conjunto de atributos se utilizaron:

categories
state
reviews_count
Stars

Nota: En el proceso, una vez que se extrajo el tipo de comida y se colocó en una nueva columna, se borro la columna “categories”

Atributos generados en el proceso.

```
8  Tip_Comida
9  DCP_FL
10 CRE_FL
11 DCP_FL_C
12 SEG_DCP_FL
13 Ind_DCP_FL_Prom
14 Ind_DCP_FL_Max
15 Prom_SEG_PEE_FL
16 Max_SEG_PEE_FL
17 PEE_FL_Cumple
```

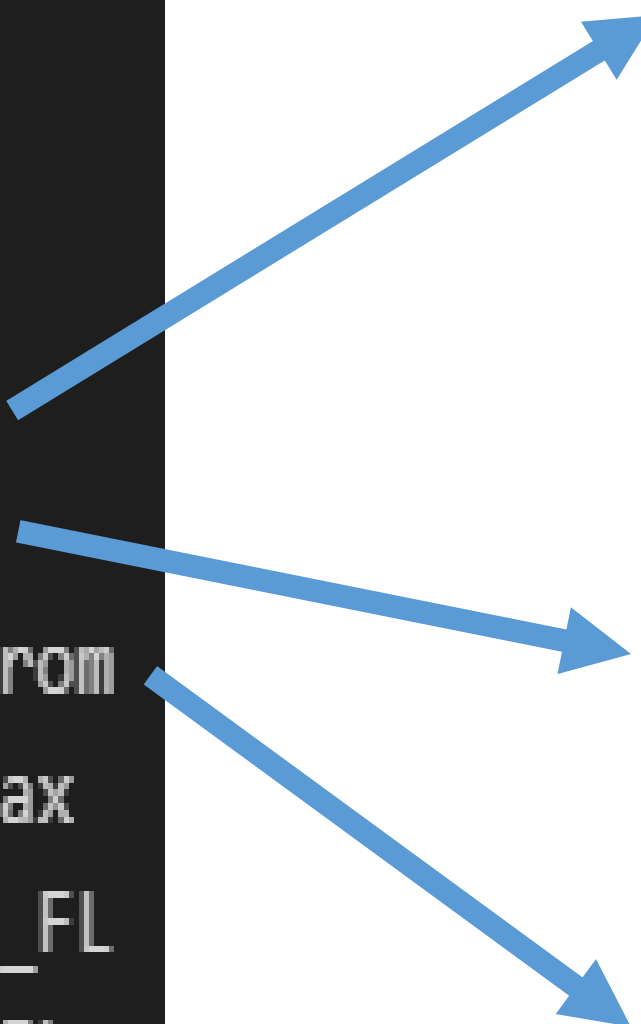


Almacena solo la palabra que define el tipo de comida a trabajar, en nuestro caso “Seafood”. Dicha palabra se busco en el atributo categories y se extrajo. Con esta nueva columna, se filtro el df por ese tipo de comida. Luego por el estado.

Se calculo el KPI del estado denominado DCP (Densidad de Consumo potencial según la Población). Es un parámetro estático del estado en su totalidad, es decir, es un único valor por el estado. Se lee: “Miles de habitantes por restaurante”. Surge de dividir (Población del estado / 1000) / (Cantidad de restaurantes en el estado).

Se calculo el KPI del estado denominado CRE (Concentración de restaurantes por territorio). Es un parámetro estático del estado en su totalidad, es decir, es un único valor por el estado. Se lee: “Cantidad de restaurantes por cada 1000 mts2 de superficie”. Surge de dividir ((Cantidad de restaurantes del tipo de comida en estudio en el estado) / ((Superficie del territorio) / 1000)).

8	Tip_Comida
9	DCP_FL
10	CRE_FL
11	DCP_FL_C
12	SEG_DCP_FL
13	Ind_DCP_FL_Prom
14	Ind_DCP_FL_Max
15	Prom_SEG_PEE_FL
16	Max_SEG_PEE_FL
17	PEE_FL_Cumple



Al estado se le calculo un DCP (Densidad de Consumo según la Población), dicho parámetro es único para el estado. Se utilizo este parámetro para generar un índice que relaciona el consumo de cada comercio con respecto al “consumo potencial del estado”, se genera entonces un “INDICE” que define a cada comercio con respecto al DCP del estado. Se calcula: $(\text{count_reviews} / \text{DCP})$. Es de hacer notar que la dist de frec de dicho nuevo parámetro es exactamente la misma que la dist de frec. De “count_reviews” pues el denominador DCP es una constante.

SEG_DCP (Segmentación de los comercios según su índice DCP), es una columna que contiene solo tres valores 1, 2 o 3. Explícitamente se refiere a la segmentación por “count_reviews”. Se segmentaron los comercios en tres tipos: Pequeños (1), Grandes (2) y Extra grandes (3).

Se tomo el atributo de DCP por comercio (DCP_C) y se calculo el promedio según el segmento al cual se corresponde cada comercio. Es decir, se calculo el promedio del índice DCP de cada comercio, por segmento.

8	Tip_Comida
9	DCP_FL
10	CRE_FL
11	DCP_FL_C
12	SEG_DCP_FL
13	Ind_DCP_FL_Prom
14	Ind_DCP_FL_Max
15	Prom_SEG_PEE_FL
16	Max_SEG_PEE_FL
17	PEE_FL_Cumple

De manera muy similar al atributo anterior. Por cada segmento de comercio, se identifico el valor máximo del índice DCP por comercio.

Promedio del nivel de percepción (stars) en el segmento al cual pertenece cada comercio

El máximo nivel de percepción (stars) en el segmento al cual pertenece cada comercio.

En esta columna solo hay dos valores 0 o 1. 1 indica que la percepción del comercio (stars) esta por encima del promedio de su segmento. 0 indica que la percepción del comercio (stars) esta por debajo del promedio de su segmento.

Rutina general y secuencial del EDA sobre cada estado.

1. Se descarga desde bigquery el archivo base filtrado por la palabra “restaurant”.
2. Ya en DataFrame, se filtra por el tipo de comida a trabajar (“Seafood”). Se crea la primera nueva columna (“Tip_Comida”).
3. Se filtra por estado (“state”).
4. Se cuenta la cantidad de restaurante (cantidad de registros)
5. Se trae el dato de superficie del estado y se calcula el CRE (Concentración de Restaurantes por Estado). Se crea una siguiente columna.
6. Se trae el dato de Población del estado y se calcula el DCP (Densidad de Consumo potencial por Estado). Se crea una siguiente columna.
7. Se genera el Índice DCP por comercio. Se crea una nueva columna (DCP_<Sigla del estado>_C).
8. Análisis de distribución de frecuencia de consumo (sobre el atributo “count_reviews”).
 1. Histograma.
 2. Calculo de estadísticos.
 3. Análisis de Outliers.
 4. Segmentación de comercios por “consumo” (usando “count_reviews”). Se crea una nueva columna (SEG_DCP _<Sigla del estado>).
 5. Calculo de estadísticos de cada segmento.
9. Análisis de distribución de comercios por “percepción” (usando “stars”) por segmentos.
 1. Histogramas por segmento de comercios.
 2. Calculo de estadísticos de percepción por cada segmento. Se crean nuevas columnas para almacenar los promedios y máximos.
 3. Se crea columna en la cual se indica por cada comercio si su percepción esta por encima del promedio o no.