

## Análisis. Actividad 5

```
#Obtengo un análisis univariado de una variable categórica en específico
table2 = freq_tbl(Micro_Retailer['312_payment_of_utilities'])
table2
```

✓ 0.0s

	312_payment_of_utilities	frequency	percentage	cumulative_perc
0	Totalmente en desacuerdo	46	0.269006	0.696970
1	Totalmente de acuerdo	9	0.052632	0.833333
2	Algo de acuerdo	5	0.029240	0.909091
3	Ni de acuerdo ni en desacuerdo	4	0.023392	0.969697
4	Algo en desacuerdo	2	0.011696	1.000000

Para la actividad decidí tomar la variable '312\_payment\_of\_utilities' para conocer cuál es la opinión acerca de los pagos de utilidades, si están de acuerdo o en desacuerdo con ellos y la frecuencia de cada una de las respuestas.

```
#Obtengo un filtro de los valores más relevantes de las variables categóricas seleccionadas más significativas
Filtro1 = table2[table2['frequency']>2]
Filtro1
```

✓ 0.0s

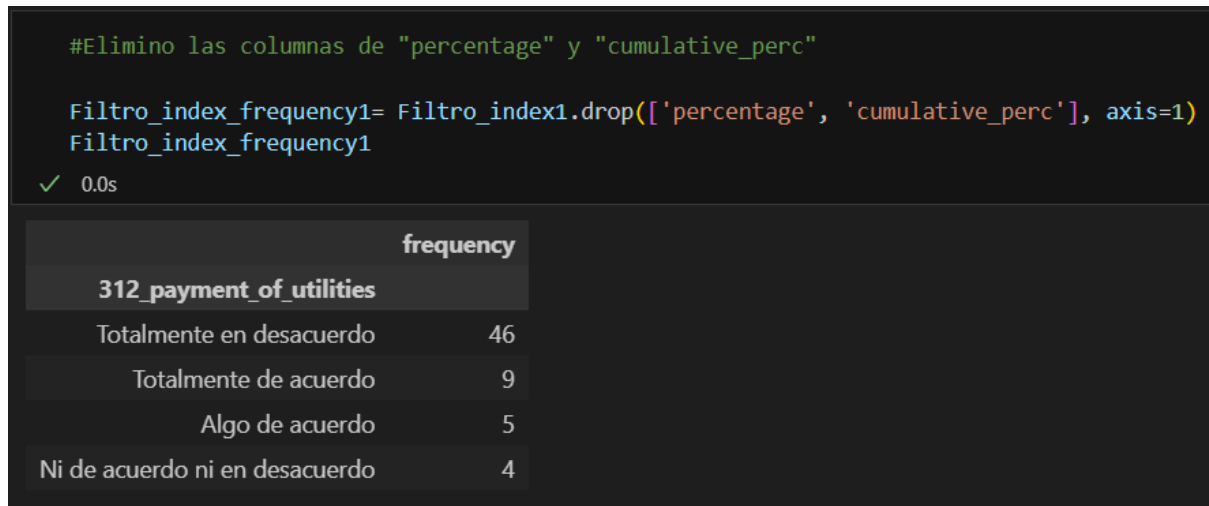
	312_payment_of_utilities	frequency	percentage	cumulative_perc
0	Totalmente en desacuerdo	46	0.269006	0.696970
1	Totalmente de acuerdo	9	0.052632	0.833333
2	Algo de acuerdo	5	0.029240	0.909091
3	Ni de acuerdo ni en desacuerdo	4	0.023392	0.969697

Filtramos los valores que consideramos más importantes de la variable categórica que seleccionamos, donde únicamente excluimos la opinión "Algo en desacuerdo" debido a que solo contaba con una frecuencia de 2 personas.

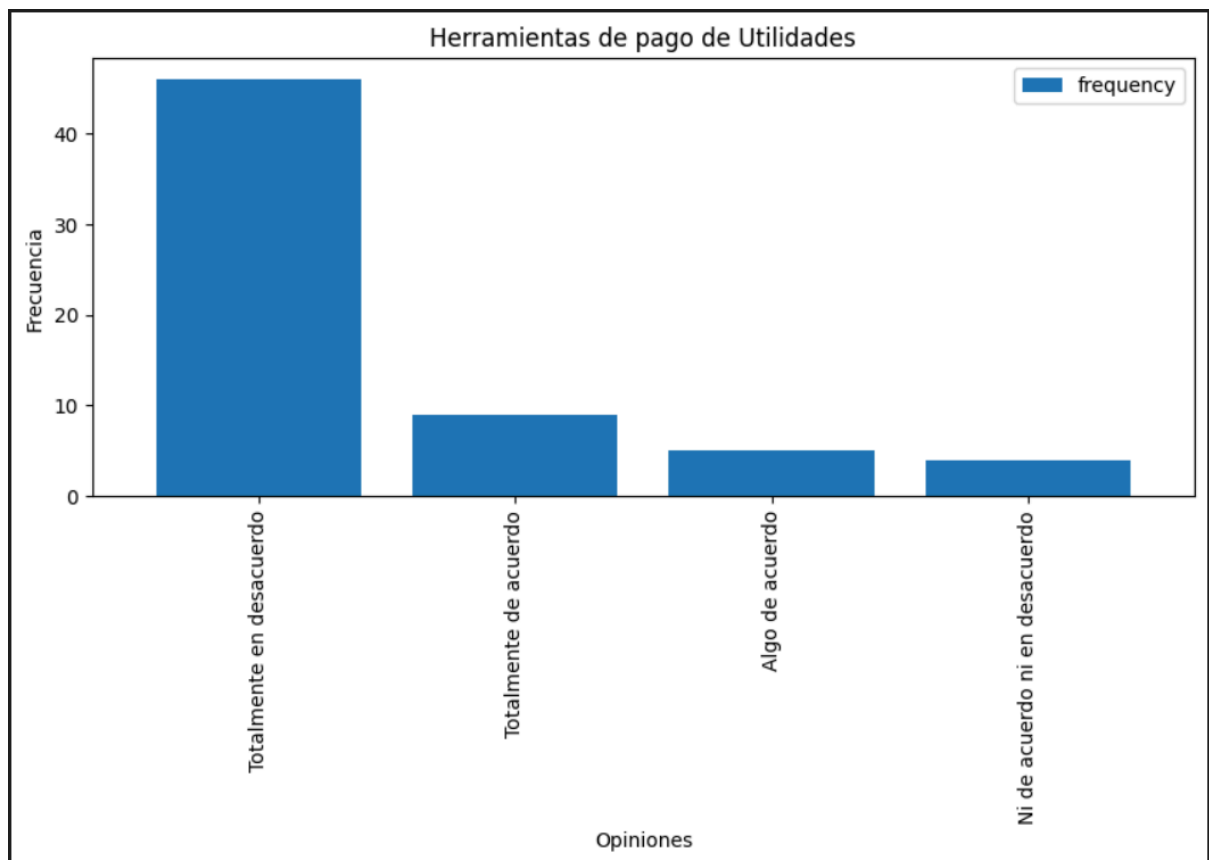
```
#Ajusto el índice de mi DataFrame
Filtro_index1=Filtro1.set_index('312_payment_of_utilities')
Filtro_index1
```

✓ 0.0s

	frequency	percentage	cumulative_perc
312_payment_of_utilities			
Totalmente en desacuerdo	46	0.269006	0.696970
Totalmente de acuerdo	9	0.052632	0.833333
Algo de acuerdo	5	0.029240	0.909091
Ni de acuerdo ni en desacuerdo	4	0.023392	0.969697

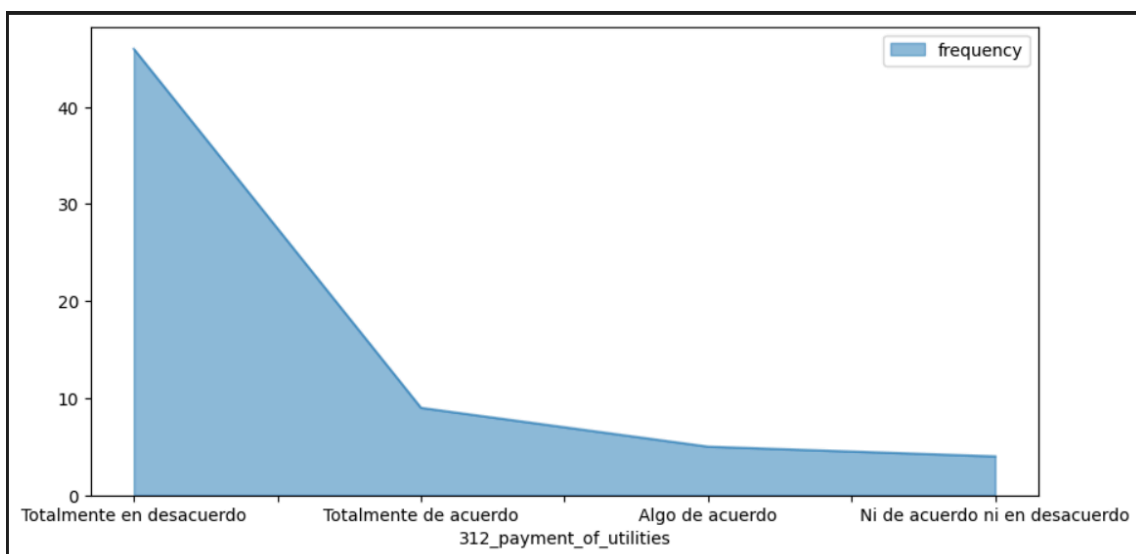
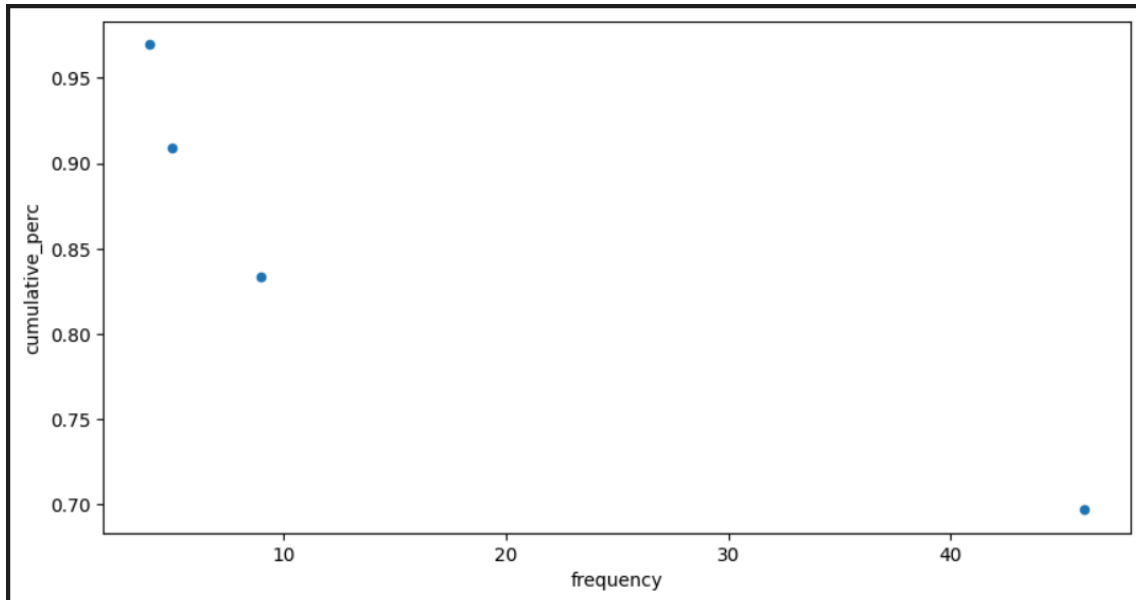


Para la siguiente gráfica quisimos ilustrar exclusivamente la frecuencia de cada una de las opiniones. Así obtuvimos que la mayoría de personas exponen que se encuentran “Totalmente en desacuerdo” con cómo se maneja el pago de utilidades. Muy por debajo de este, siendo menos de 10 personas las que opinaron, dicen estar “Totalmente de acuerdo”, seguido de 5 diciendo “Algo en desacuerdo” y 4 exclamando que no están “Ni de acuerdo ni en desacuerdo”.



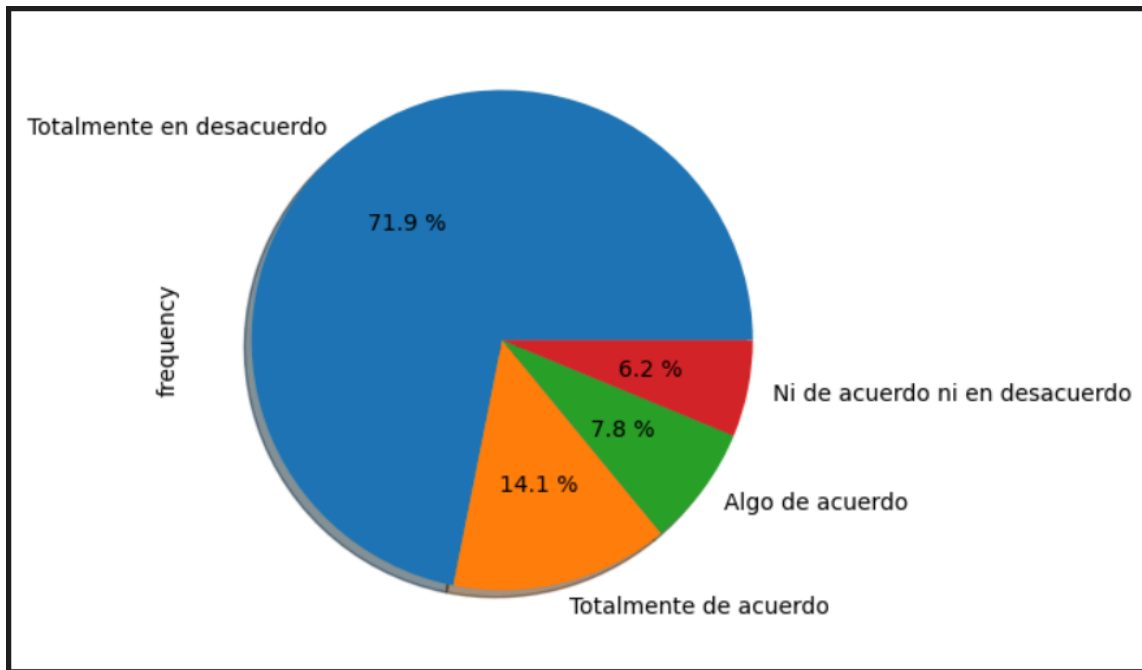
Ahora, ilustramos un gráfico de dispersión que muestra la relación entre la “frecuencia” y el “porcentaje acumulado” para las diferentes respuestas de la encuesta. Gracias a ello, se muestra que la mayor parte de la frecuencia (46) se encuentra en la categoría “Totalmente

en desacuerdo", con un porcentaje acumulado de aproximadamente 0.70 o 70%, y a medida que disminuye la frecuencia, el porcentaje acumulado aumenta, indicando que las respuestas con menor frecuencia contribuyen más al porcentaje acumulado, y el punto en el extremo derecho del gráfico representa la frecuencia baja (de 4 personas) pero con un porcentaje acumulado alto.

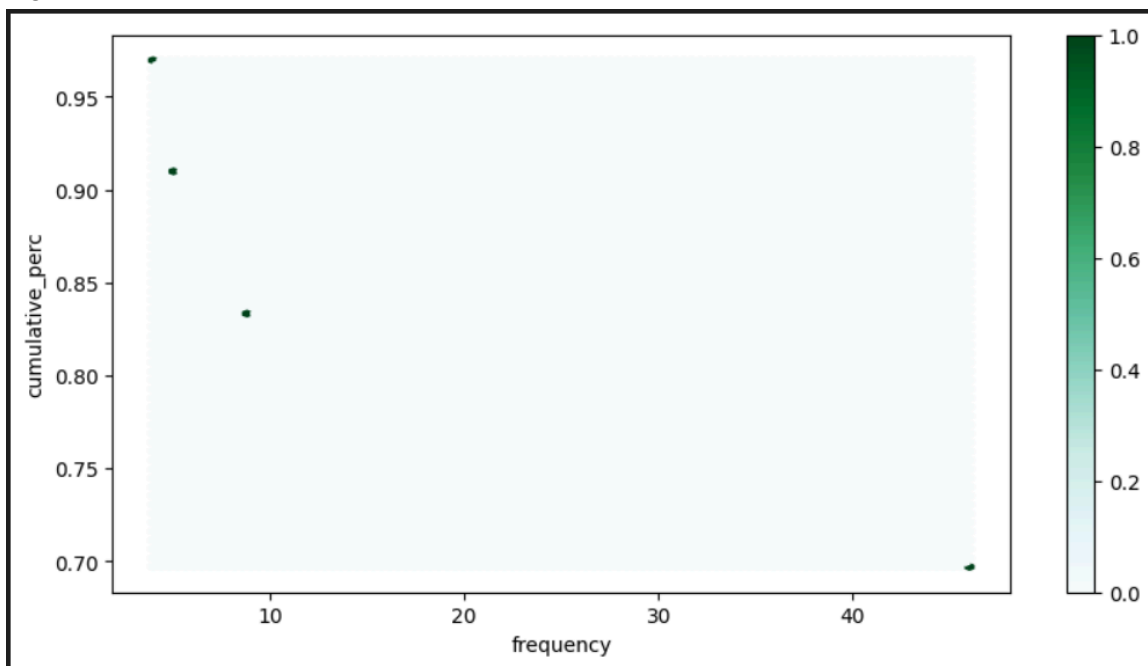


Ahora, el anterior es un gráfico de área, que es similar a un gráfico de líneas, pero el área debajo de la línea se rellena, lo que enfatiza la magnitud de los valores en relación con el eje X. Este ilustra lo anteriormente mencionado, pero se ilustra de una manera distinta, facilitando la visualización de cómo se distribuyen las frecuencias de las respuestas en comparación entre sí.

Por otra parte, ilustramos un gráfico de pastel con los mismos datos de frecuencia, pero en este caso se pueden apreciar los porcentajes exactos de cada una de las opiniones, así como a cuál corresponde cada una.



Por último , realizamos un gráfico hexagonal, el cual muestra en el eje X la Frecuencia de las respuestas y en el eje Y el porcentaje de acumulado de respuestas. Lo que hace el gráfico hexagonal es agrupar los datos en celdas hexagonales en lugar de puntos individuales como en un gráfico de dispersión, donde cada celda representa una región de valores en la gráfica, y su color indica la cantidad de datos o puntos que caen dentro de esa región.



Este sirve como una alternativa al gráfico de dispersión que facilita la visualización de la densidad de los datos, más porque las frecuencias son limitadas (46, 9, 5, 4) y los porcentajes acumulados tienen un rango específico de 0.70 a 0.97. Así, las celdas hexagonales se concentraron en las zonas donde los valores de frecuencia y porcentaje acumulado están presentes, haciendo más evidente la relación entre estos dos parámetros en la distribución de las respuestas de la encuesta, no obstante, el gráfico sería más útil para casos donde los conjuntos de datos son más grandes o dispersos.