

# Visualización de la Información

## Tarea 4

### Resumen

En este reporte se abordarán actividades tanto de reducción de dimensionalidades y graficación en el nuevo subespacio, así mismo como graficación de imágenes y aplicación de interpolación para comparar calidad de las mismas.

Como **objetivo** tenemos que la visualización es muy importante para explicar nuestros datos, sin embargo el hecho de que las visualizaciones gráficas únicamente pueden llegar hasta 3 dimensiones nos limitan a tener que aplicar cierto preprocesamiento a nuestros datos sin llegar a afectar la forma en la que se ven, por lo que implementaremos algoritmos de reducción de dimensionalidades para poder graficar los datos.

Por otro lado las visualizaciones así como nuestras imágenes, pierden calidad y nitidez cada que nos acercamos (zoom) computacionalmente, por lo que estaremos aprendiendo el cómo poder incrementar esta calidad para poder observar de mejor manera nuestra imagen y así lograr obtener los mayores detalles posibles.

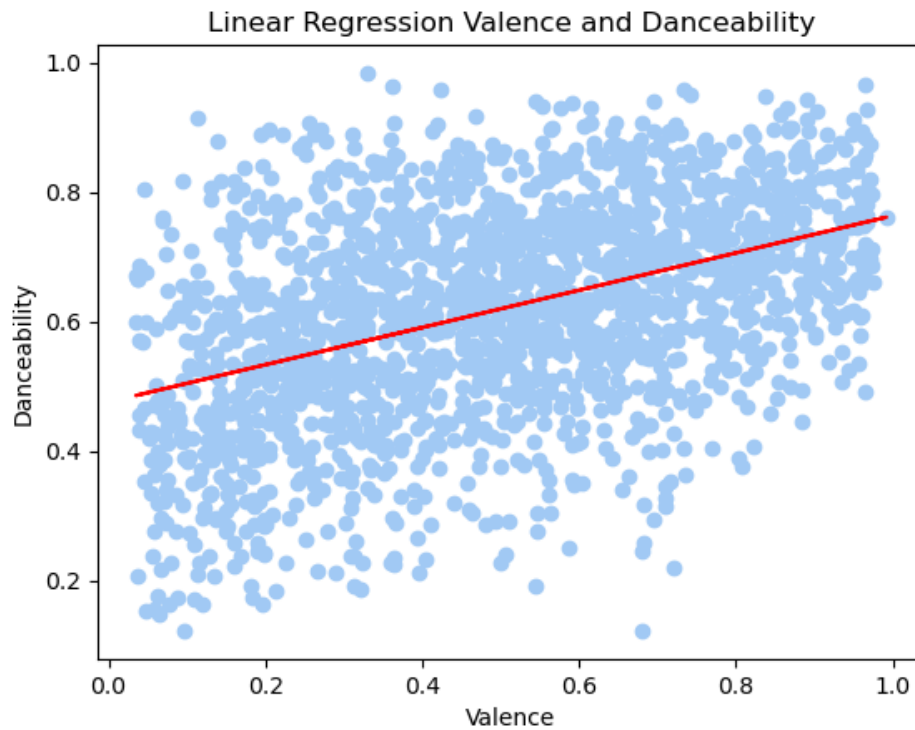
### Actividad 1

Utilice el conjunto de datos [spotify.csv](#) para análisis. Los datos contienen métricas numéricas generadas por Spotify que miden la capacidad de baile (danceability), el estado de ánimo (mood), la vivacidad (liveness), etc. de las canciones. Los datos también contienen el título y el artista de las canciones.

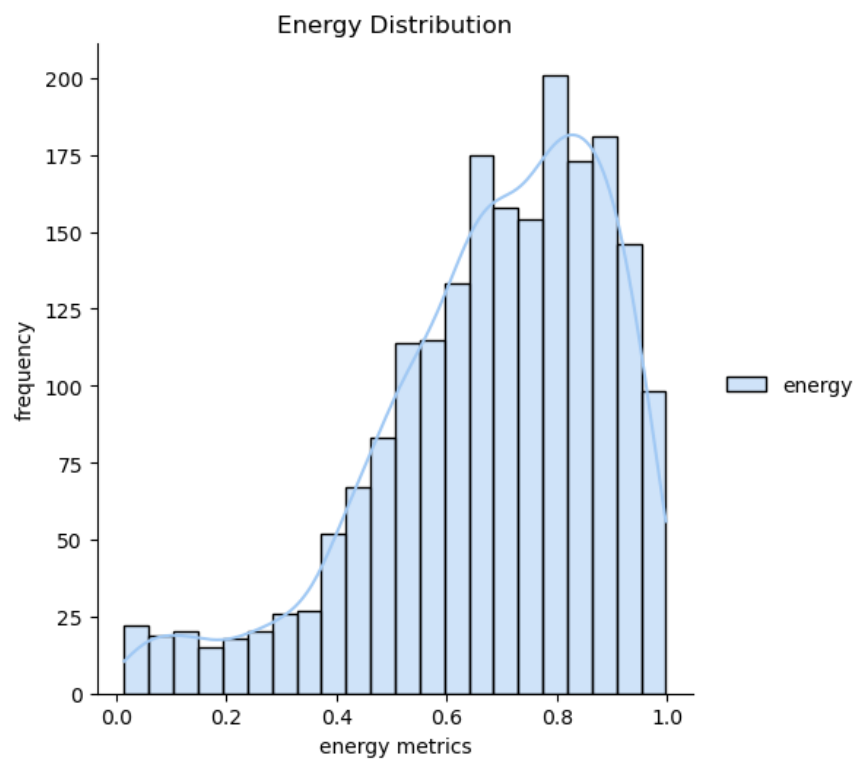
### Dataset

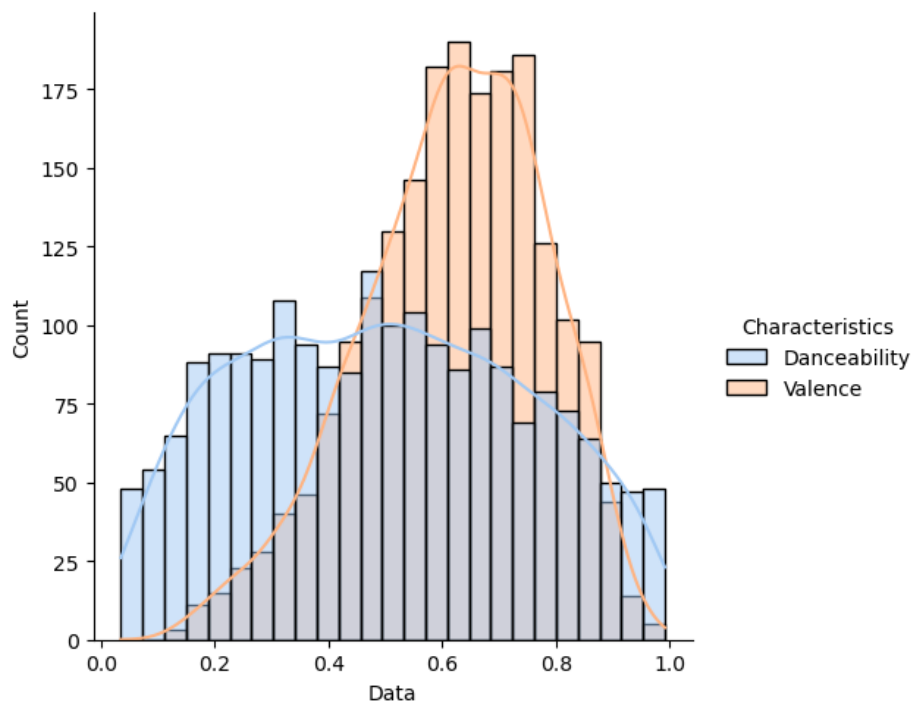
	Unnamed: 0	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence	target	song_title	artist
0	0	0.0102	0.833	204600	0.434	0.021900	2	0.1650	-8.795	1	0.4310	150.062	4.0	0.286	1	Mask Off	Future
1	1	0.1990	0.743	326933	0.359	0.006110	1	0.1370	-10.401	1	0.0794	160.083	4.0	0.588	1	Redbone	Childish Gambino
2	2	0.0344	0.838	185707	0.412	0.000234	2	0.1590	-7.148	1	0.2890	75.044	4.0	0.173	1	Xanny Family	Future
3	3	0.6040	0.494	199413	0.338	0.510000	5	0.0922	-15.236	1	0.0261	86.468	4.0	0.230	1	Master Of None	Beach House
4	4	0.1800	0.678	392893	0.561	0.512000	5	0.4390	-11.648	0	0.0694	174.004	4.0	0.904	1	Parallel Lines	Junior Boys

## Scatterplots de Valencia y Bailabilidad



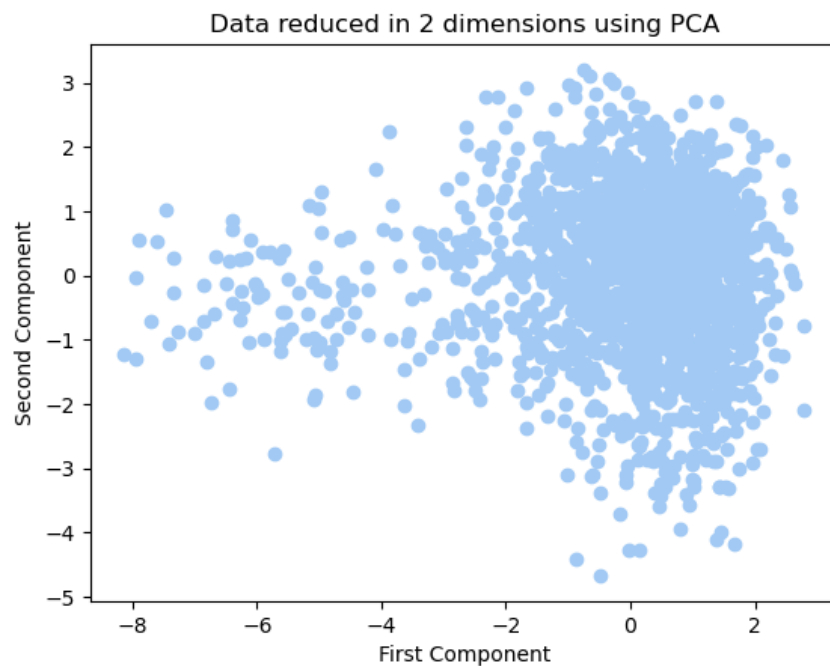
## Distribución en Función tanto como de Energía, como de Valencia y Bailabilidad



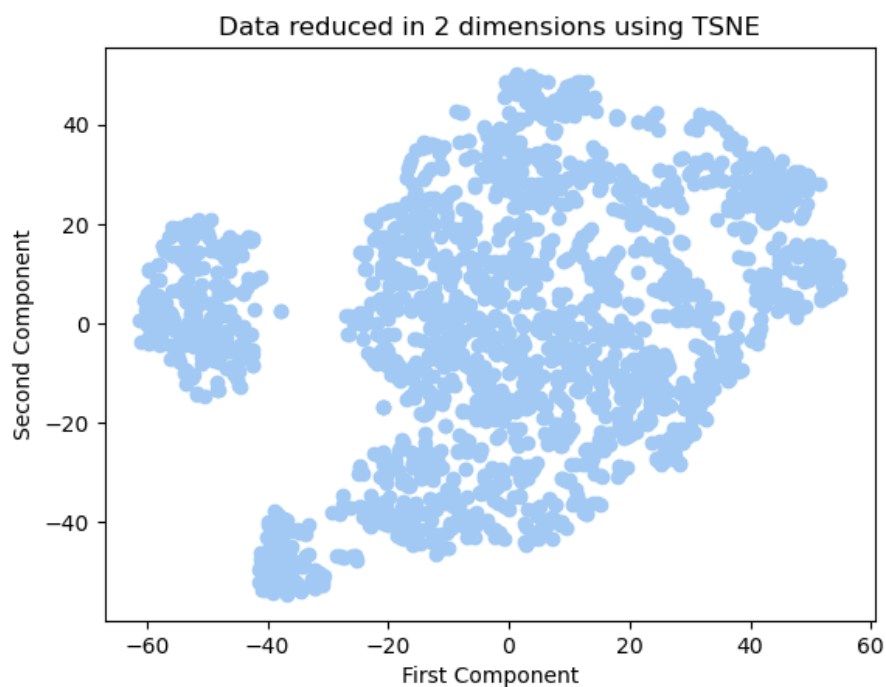


A continuación, genere una lista de rasgos "elegidos" y use el **análisis de componentes principales** (PCA) para reducir las dimensiones de esa lista. Cuanto menor sea la distancia entre dos canciones, mayores serán las similitudes en sus rasgos.

Plot del Dataset Reducido por PCA



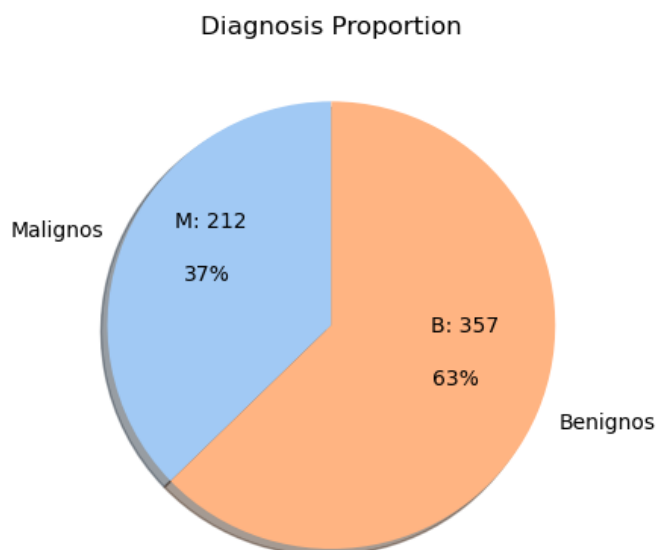
## Plot del Dataset Reducido por t-SNE



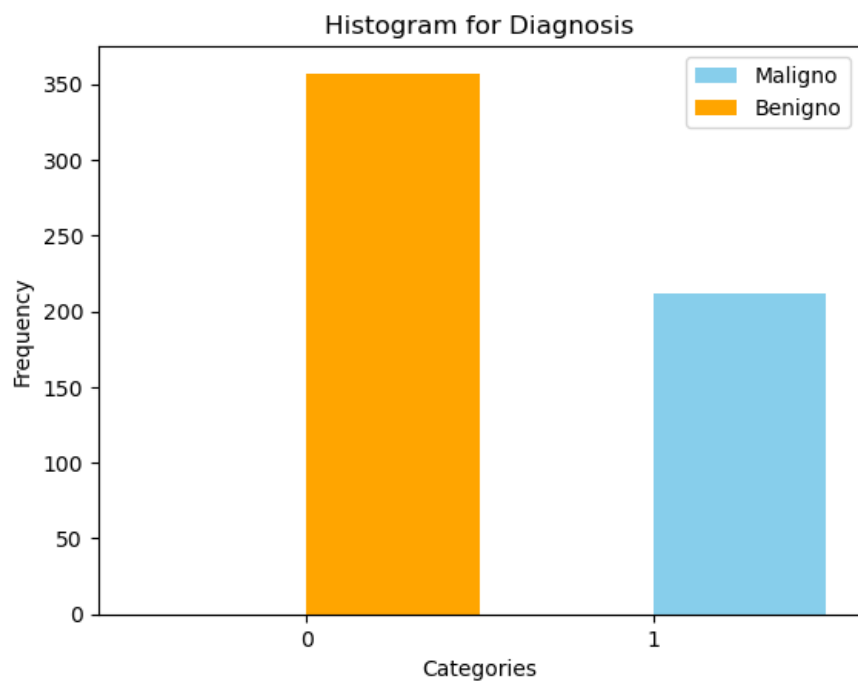
## Actividad 2

Utilice el conjunto de datos [cancer.csv](#) para el análisis. Los datos contienen métricas numéricas sobre análisis de cáncer de mama.

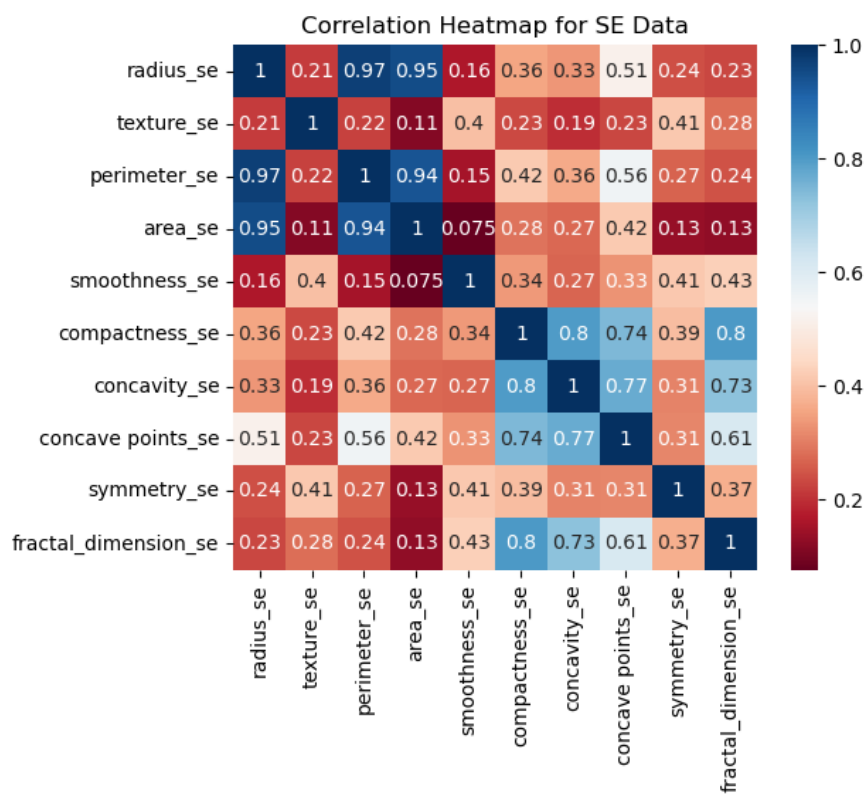
### Plot de Benignos y Malignos



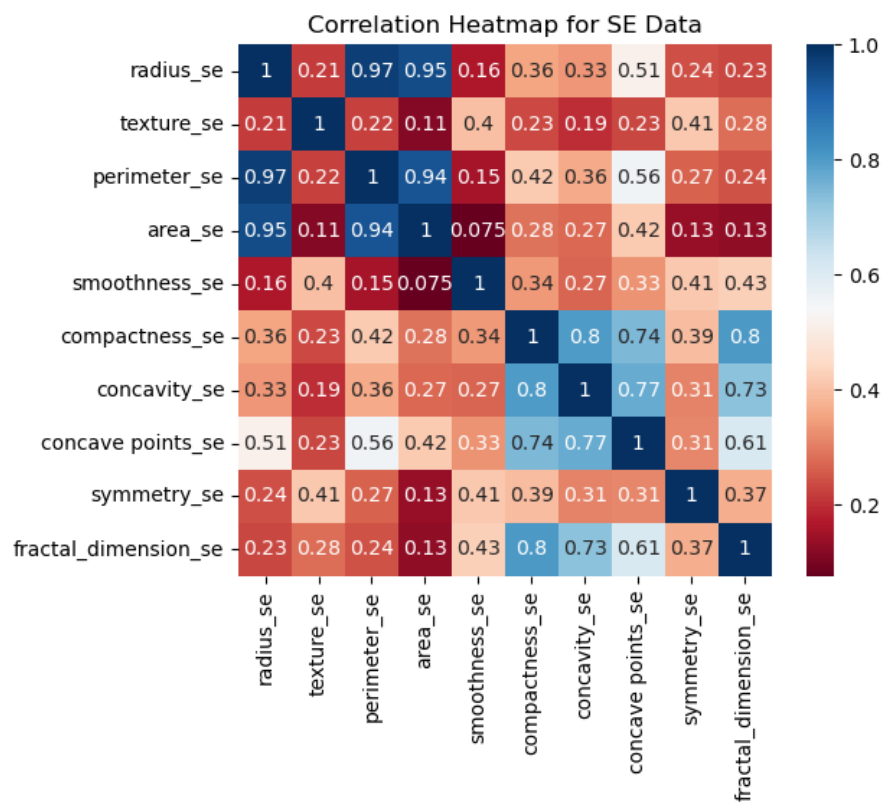
## Plot de Benignos y Malignos Etiquetados con Números



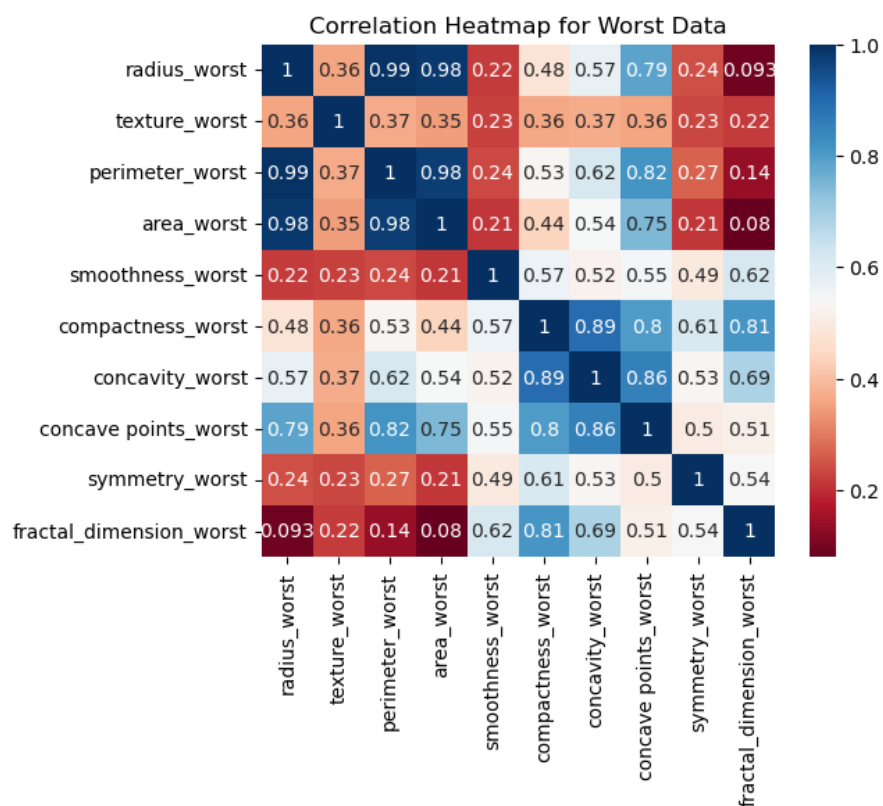
## Correlation Matrix Mean



## Correlation Matrix SE

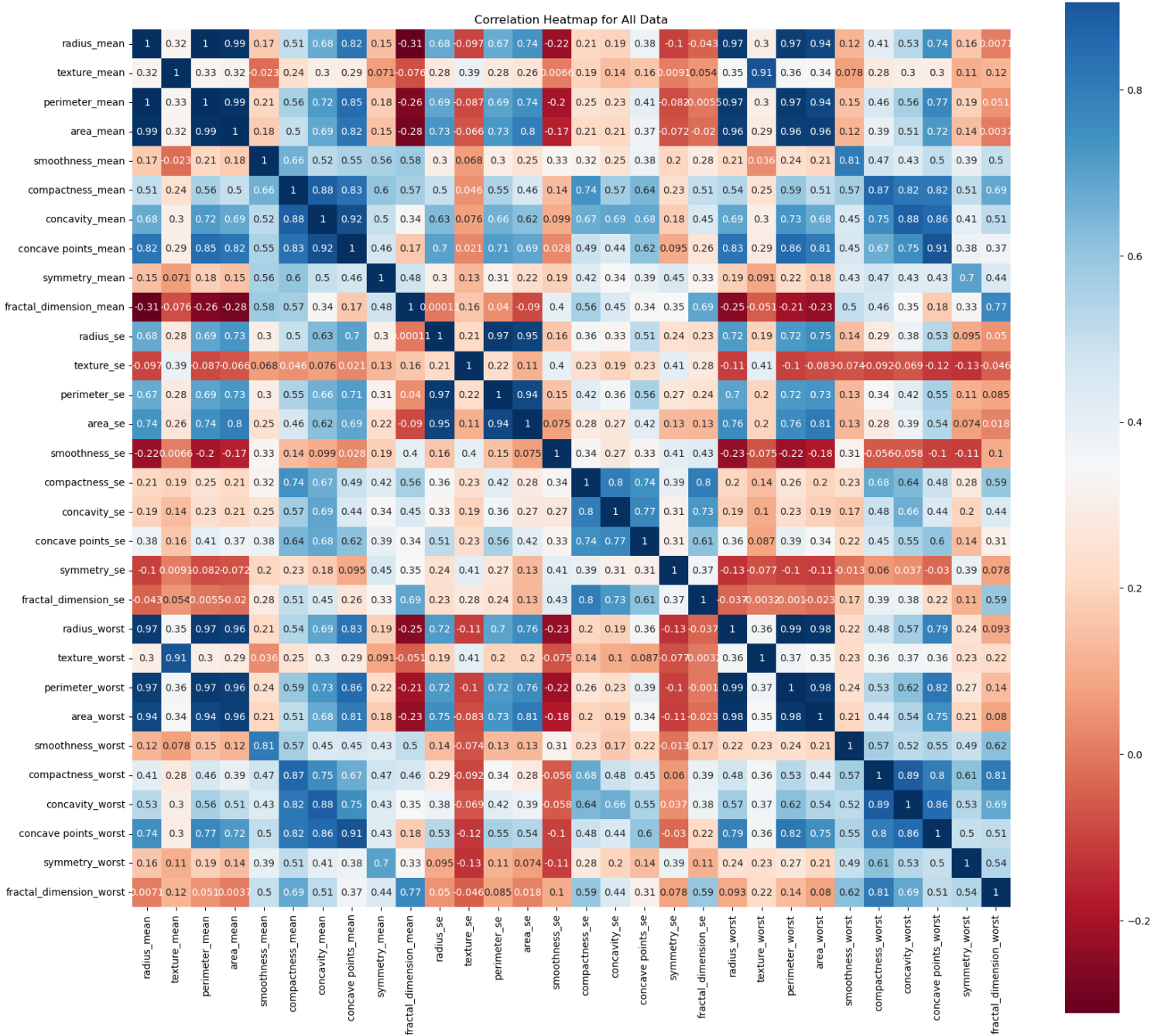


## Correlation Matrix Worst



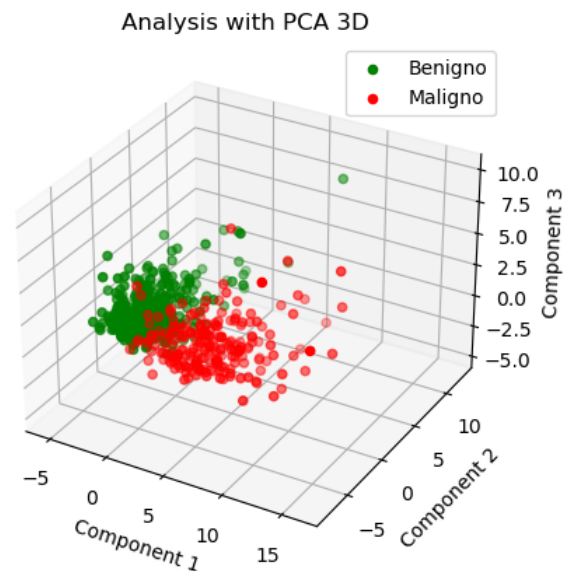
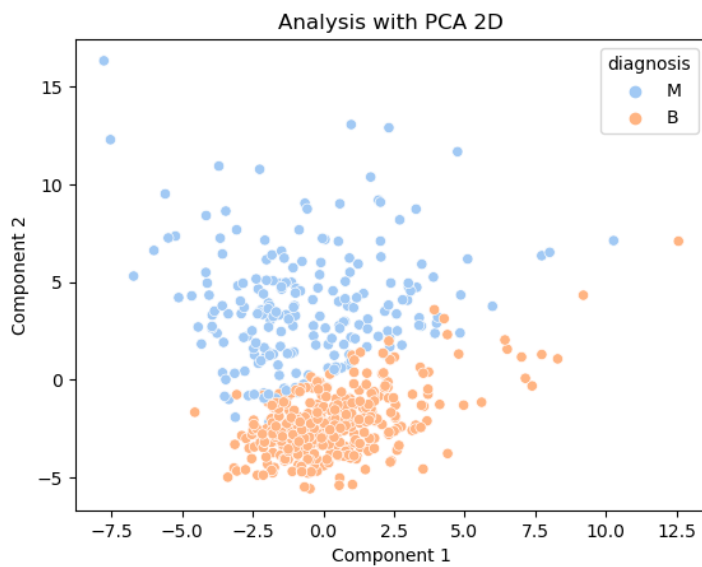
# Correlation Matrix

Correlation Heatmap for All Data

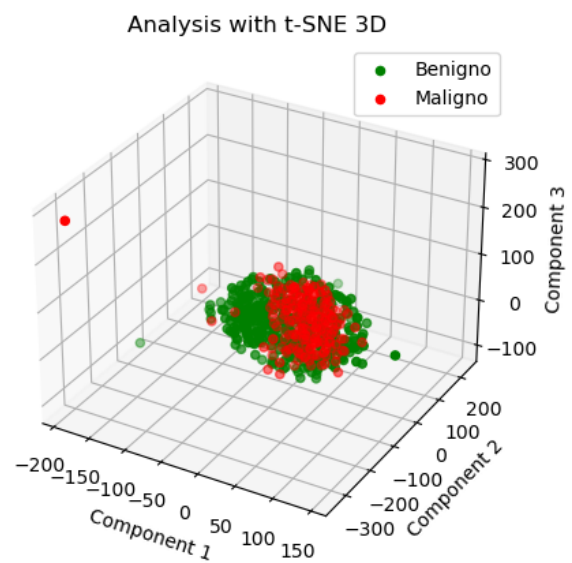
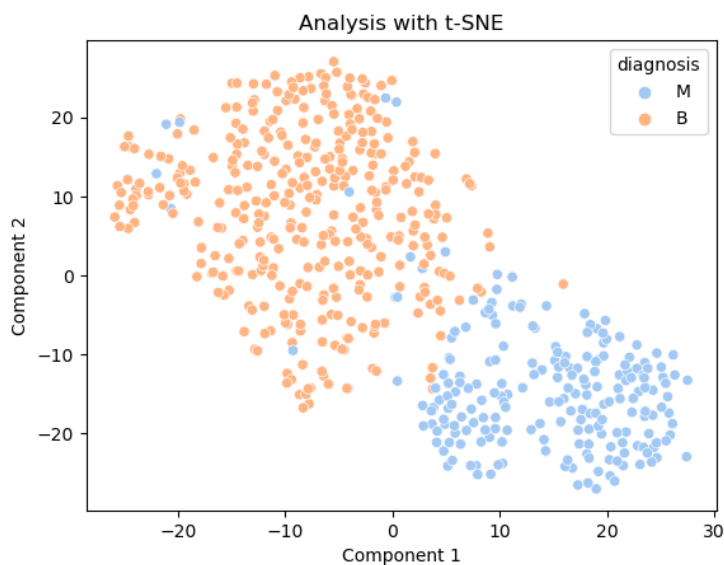


A partir de las correlaciones observadas, intente eliminar alguna de las características para ver si afecta a la clasificación. En caso de no afectar, sería positivo eliminarlas, ya que reduce el número de características a tratar, lo que es beneficioso de cara al coste computacional. Por ejemplo: radio (radio), perímetro (perimeter)) y área (area), Concavidad (Concavity), puntos cóncavos (concave points) y compacidad (compactness).

## Análisis de Benignos y Malignos con PCA 2D y 3D



## Análisis de Benignos y Malignos con t-SNE





## Actividad 3

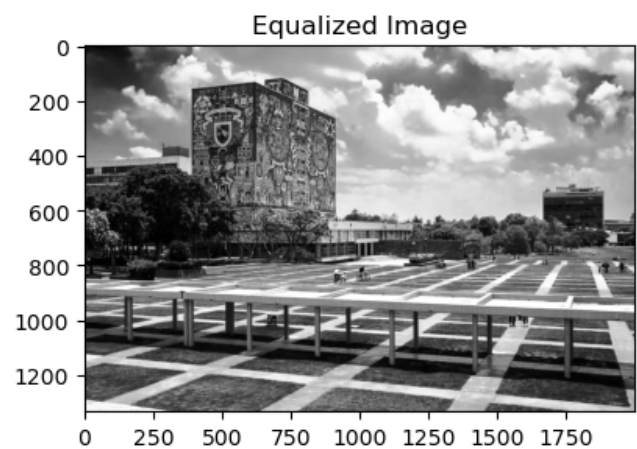
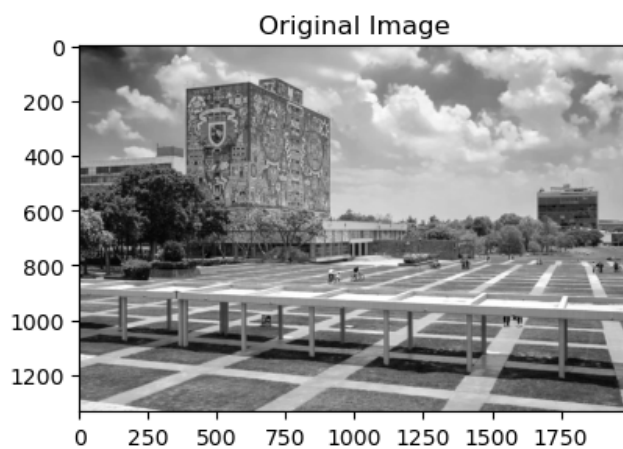
Considere la siguiente imagen, la cual se puede visualizar en el **archivo adjunto**:

[unam.png](#)

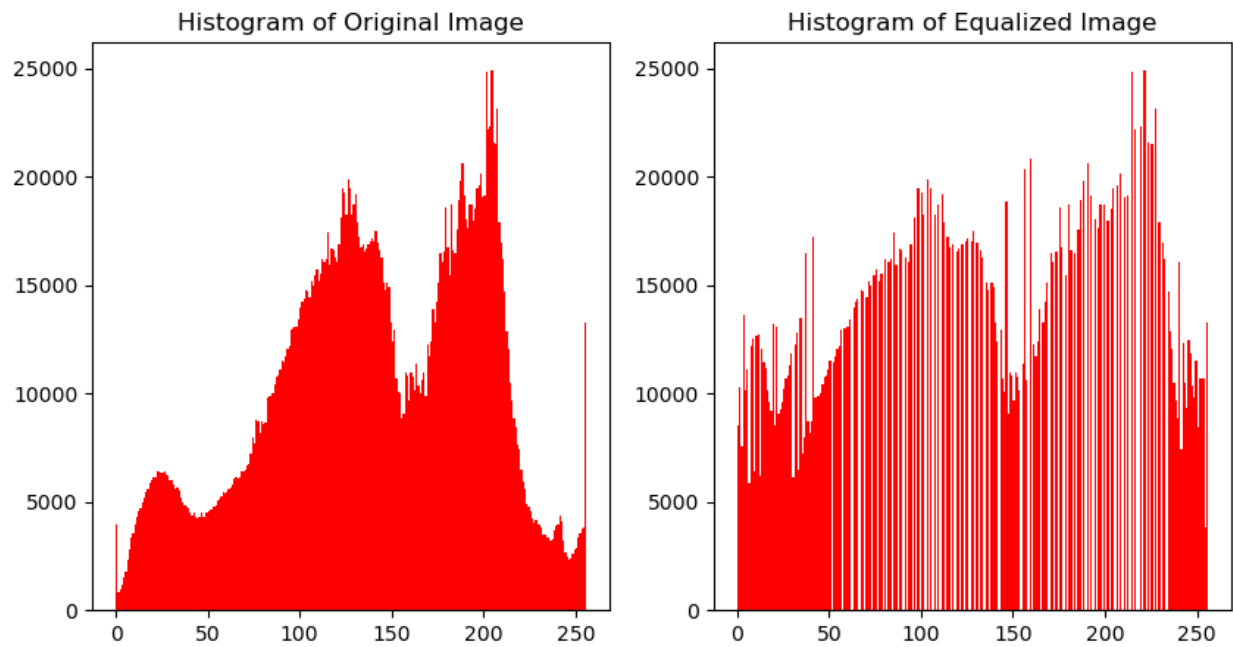


## Ecualización de Histograma y Descripción

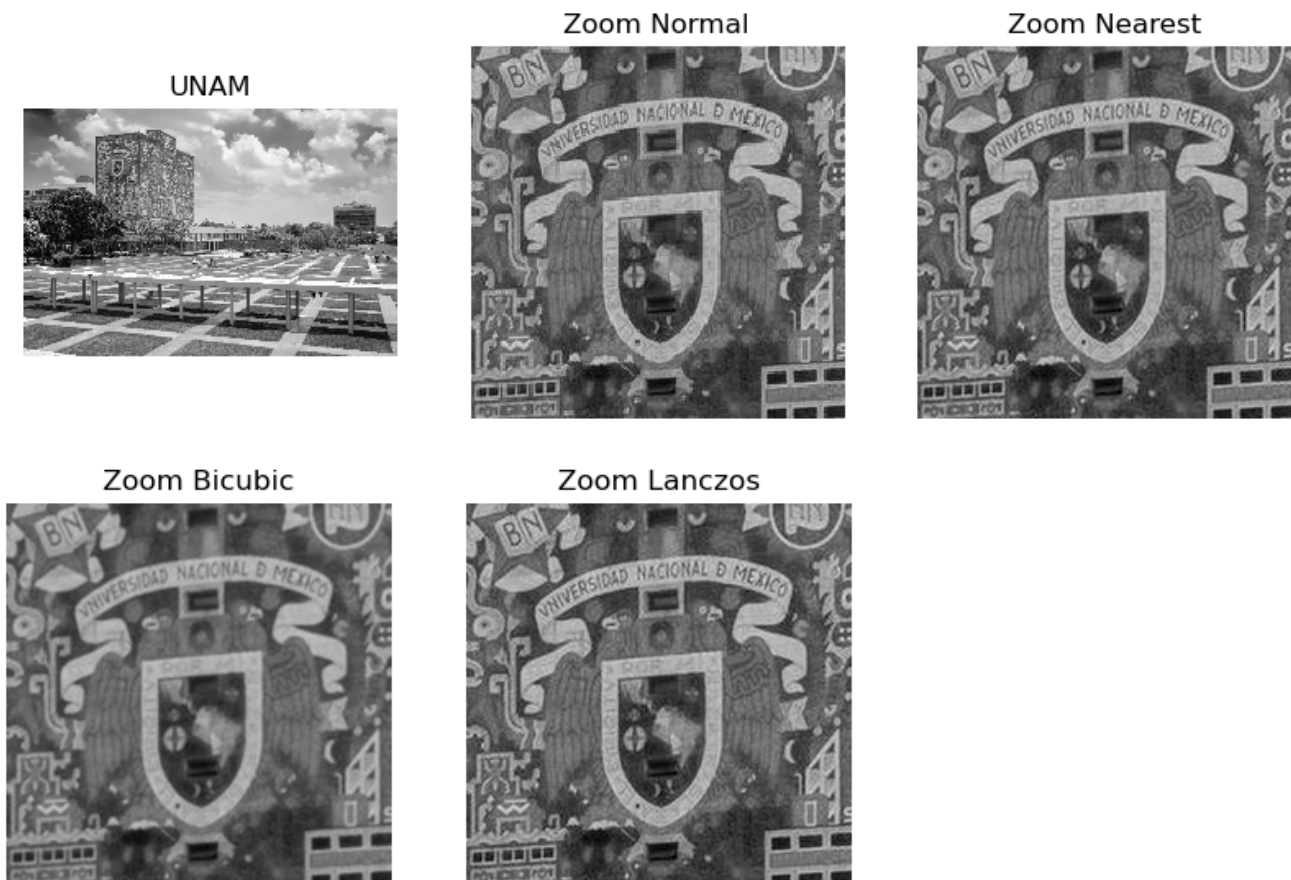
Imágenes ecualizadas




## Histograma de la Ecualización



## Regiones con Zoom y Comparación



- 
- Nearest : El zoom en la foto nearest con respecto al original en el caso de esta imagen parecen ser casi idénticas, sin embargo en nearest parecen estar los bordes más marcados por pixeles que por difuminaciones.
  - Bicubic : Esta interpolación deja ver bastante claro que incluso la imagen normal se ve mejor y más nítida que aplicando *Nearest*, pues las letras aparentan tener un filtro medio borroso.
  - Lanczos : Este tipo de interpolación nos incrementa las texturas y la nitidez en las imágenes debido a que también parece que incrementa en un factor muy pequeño el ruido de la foto.

## Conclusión

La tarea logró realizarse de manera satisfactoria y única. Se lograron captar los conocimientos e ideas sobre la visualización y sobre el procesamiento de imágenes. Así mismo que solo es posible graficar para 3 dimensiones lo cuál el desarrollo del algoritmo PCA si bien fue tedioso, es uno de los algoritmos más famosos y populares.

Posteriormente la práctica 3 me otorgó todos los conocimientos que se necesitaban para esta tarea, por lo que la parte de la visualización de las imágenes igual resultó satisfactoria.

