BANANA PROBLEM STATEMENT

About Dimensionality reduction:

This lab explores the application of machine learning techniques to classify songs into two genres: 'Hip-Hop' and 'Rock'. We use Decision Trees and Logistic Regression for classification, along with data preprocessing, feature scaling, and dimensionality reduction using Principal Component Analysis (PCA).

Dimensionality reduction is a technique used to reduce the number of features in a dataset while retaining as much relevant information as possible. It simplifies the dataset by transforming it into a lower-dimensional space. This process can improve model performance, reduce computational cost, help visualize high-dimensional data, and reduce overfitting.

Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction in datasets. It transforms the original data into a new coordinate system where the axes (called principal components) represent the directions of maximum variance in the data. By projecting the data onto these new axes, PCA simplifies the dataset while preserving as much of the original information as possible.
This helps in reducing complexity and increases visualization.

How PCA Works:

- **Centering the Data:** Subtract the mean of each feature to center the data around the origin.
- **Computing the Covariance Matrix:** Calculate how features vary with respect to each other.
- **Finding Principal Components:** Compute the eigenvectors (principal components) and eigenvalues of the covariance matrix. The eigenvectors represent directions of maximum variance, and the eigenvalues indicate the amount of variance in those directions.
- **Transforming the Data:** Project the original data onto the principal components to obtain the reduced feature set.

About the dataset:

The dataset for this classification task consists of musical features and genre labels for various tracks. The primary goal is to classify each song into one of the two genres ('Hip-Hop' or 'Rock') based on these features

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	track_id	bit_rate	comments	composer	date_created	date_recorded	duration	favorites	genre_top	genres	genres_all	information	interest	language_	license	listens	lyricist	number	publisher	tags	title		
2	135	256000	1		#########	##########	837	0	Rock	[45, 58]	[58, 12, 45]		2484	en	Attribution	1832		0		[]	Father's Day		
3	136	256000	1		#########	##########	509	0	Rock	[45, 58]	[58, 12, 45]		1948	en	Attribution	1498		0		[]	Peel Back The Mountain Sky		
4	151	192000	0		#########		192	0	Rock	[25]	[25, 12]		701	en	Attribution	148		4		[]	Untitled 04		
5	152	192000	0		#########		193	0	Rock	[25]	[25, 12]		637	en	Attribution	98		11		[]	Untitled 11		
6	153	256000	0	Arc and Se	#########	##########	405	5	Rock	[26]	[26, 12]		354	en	Attribution	424		2		[]	Hundred-Year Flood		
7	154	256000	0	Arc and Se	#########	##########	319	1	Rock	[26]	[26, 12]		242	en	Attribution	205		4		[]	Squares And Circles		
8	155	192000	0		#########	##########	756	1	Rock	[26]	[26, 12]		268	en	Attribution	197		0		[]	Maps of the Stars Homes		
9	169	192000	0	James Squ	#########	##########	144	1	Rock	[25]	[25, 12]		815	en	Attribution	270		1		[]	Boss of Goth		
10	170	192000	0		#########	##########	181	0	Rock	[25]	[25, 12]		468	en	Attribution	122		2		[]	Industry Standard Massacre		
11	171	320000	0		#########	##########	80	0	Rock	[25]	[25, 12]	<p>Unrele	582	en	Attribution	242		2		[]	Marching as Technitions		
12	172	192000	0	James Squ	#########	##########	206	0	Rock	[25]	[25, 12]		468	en	Attribution	90		3		[]	I Can't Stand You Up-right		