Spotify Project 4
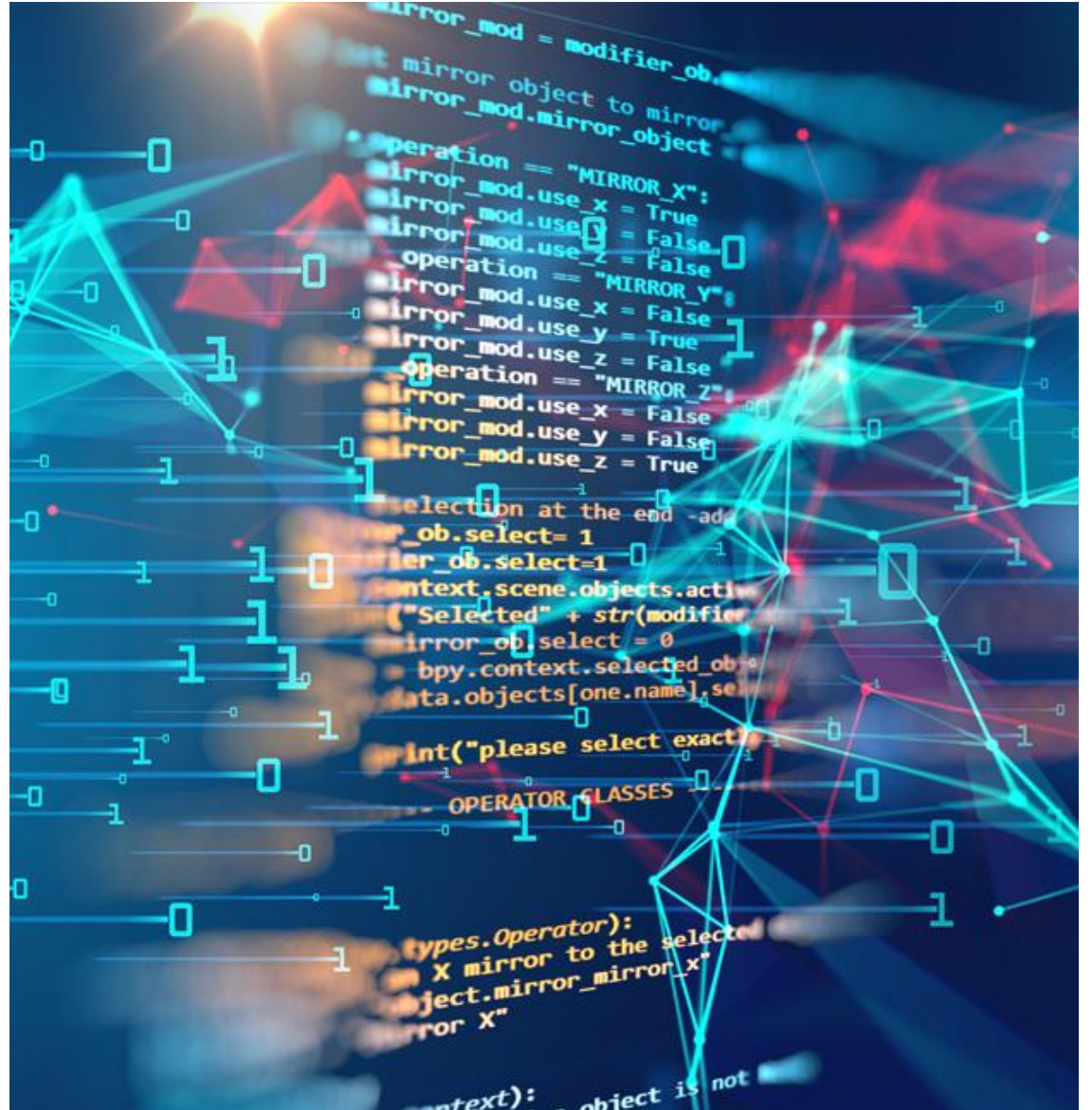
Tehreem uzma, Christopher Levine,
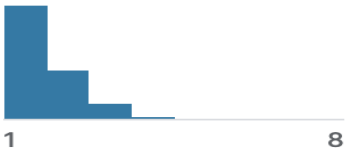Justin Shilling, Abanoub Malek,
Mark Habeb

Project Overview

- Our goal was to analyze a dataset of the most streamed songs on Spotify to determine the factors influencing song popularity.

- We utilized machine learning models, to predict song success

DATASET

https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023

△ track_name	△ artist(s)_name	# artist_count	# released_year	# released_month
Name of the song	Name of the artist(s) of the song	Number of artists contributing to the song	Year when the song was released	Month when the song was released
943 unique values	Taylor Swift 4% The Weeknd 2% Other (897) 94%	1 — 8	1930 — 2023	1 — 12
Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7
LALA	Myke Towers	1	2023	3
vampire	Olivia Rodrigo	1	2023	6
Cruel Summer	Taylor Swift	1	2019	8
WHERE SHE GOES	Bad Bunny	1	2023	5
Sprinter	Dave, Central Cee	2	2023	6
Ella Baila Sola	Eslabon Armado, Peso Pluma	2	2023	3
Columbia	Quevedo	1	2023	7

DATASET OVERVIEW

| spotify_df	
track_name	VARCHAR,
artist_name	VARCHAR,
artist_count	INT,
released_year	INT,
released_month	INT,
released_day	INT,
streams	BIGINT,
bpm	INT,
key	VARCHAR,
mode	VARCHAR,
danceability_pct	INT,
valence_pct	INT,
energy_pct	INT,
acousticness_pct	INT,
instrumentalness_pct	INT,
liveness_pct	INT,
speechiness_pct	INT
stream_category	VARCHAR
song_status	INT

kaggle
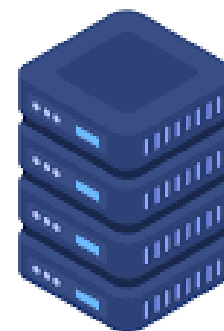
```
['Track Name', 'Artist Name', 'Artist Count', 'Released Year',
 'Released Month', 'Released Day', 'Spotify Chart Rank', 'Streams',
 'Apple Chart Rank', 'BPM', 'Mode', 'Danceability Pct', 'Valence Pct',
 'Energy Pct', 'Acousticness Pct', 'Instrumentalness Pct',
 'Liveness Pct', 'Speechiness Pct', 'Stream Category'],
```
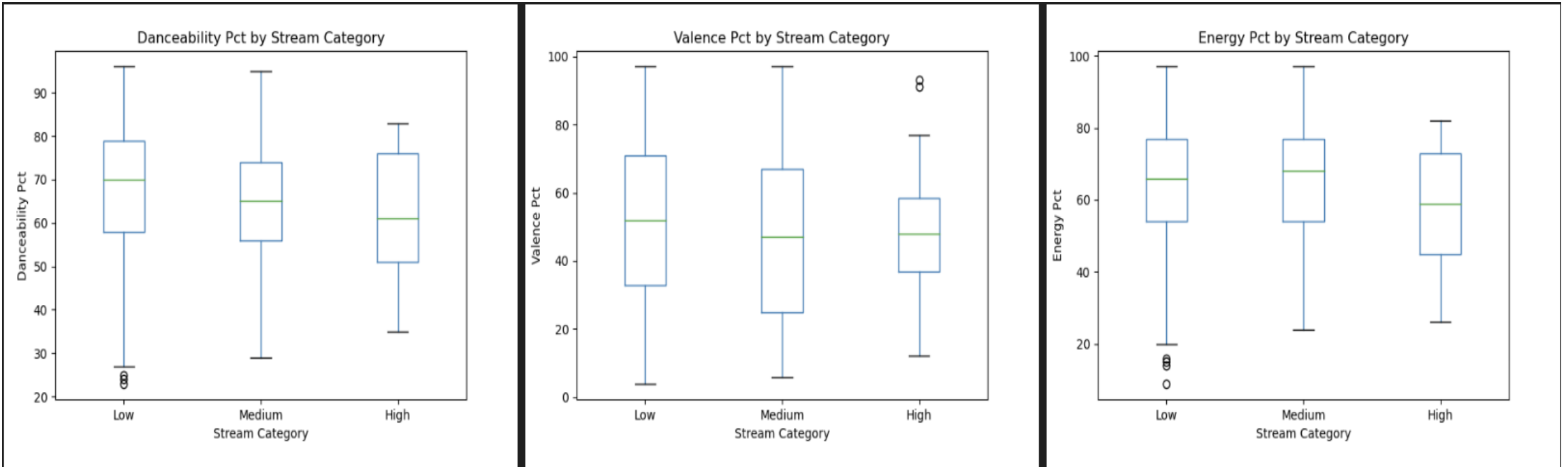
FINDING A PATTERN

- **danceability_%**: *Percentage indicating how suitable the song is for dancing*

- **valence_%**: *Positivity of the song's musical content*

- **energy_%**: *Perceived energy level of the song*

- **acousticness_%**: *Amount of acoustic sound in the song*

- **instrumentalness_%**: *Amount of instrumental content in the song*

- **liveness_%**: *Presence of live performance elements*

- **speechiness_%**: *Amount of spoken words in the song*
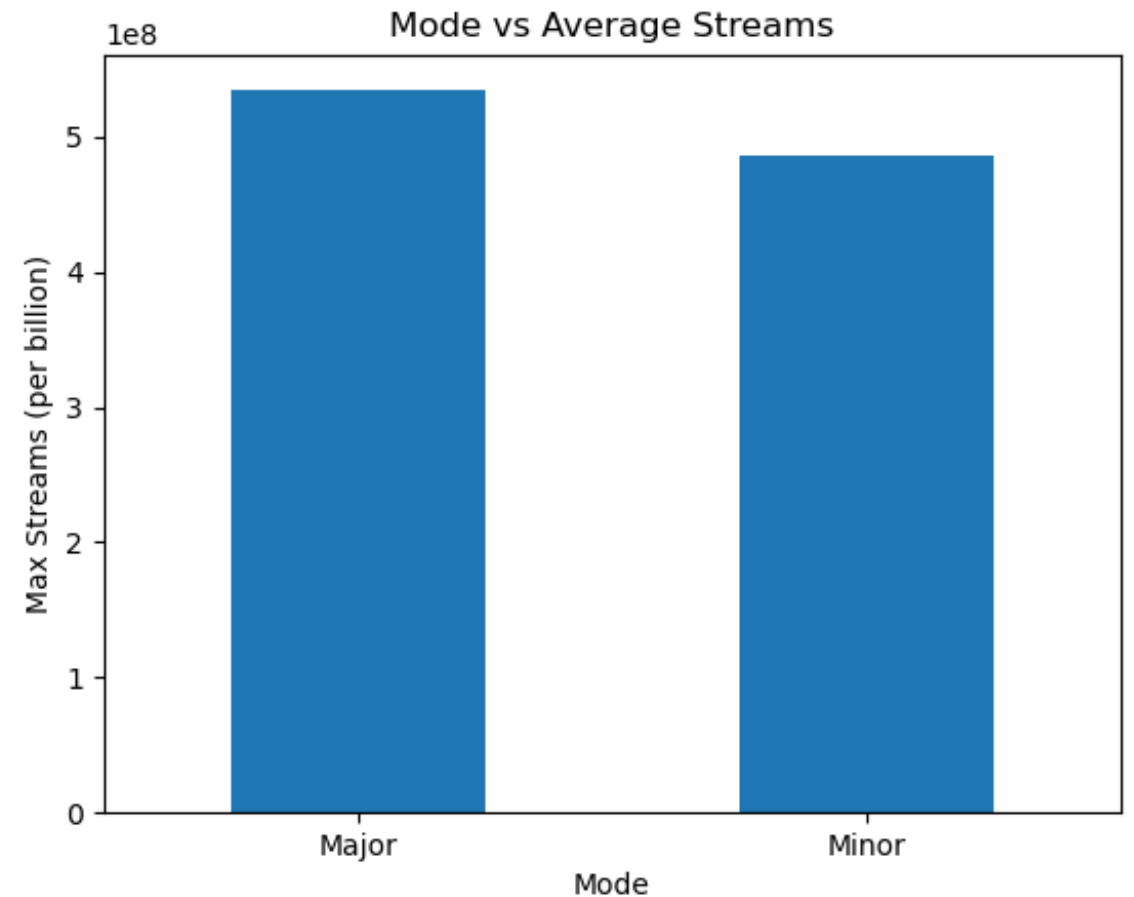
FINDING A PATTERN
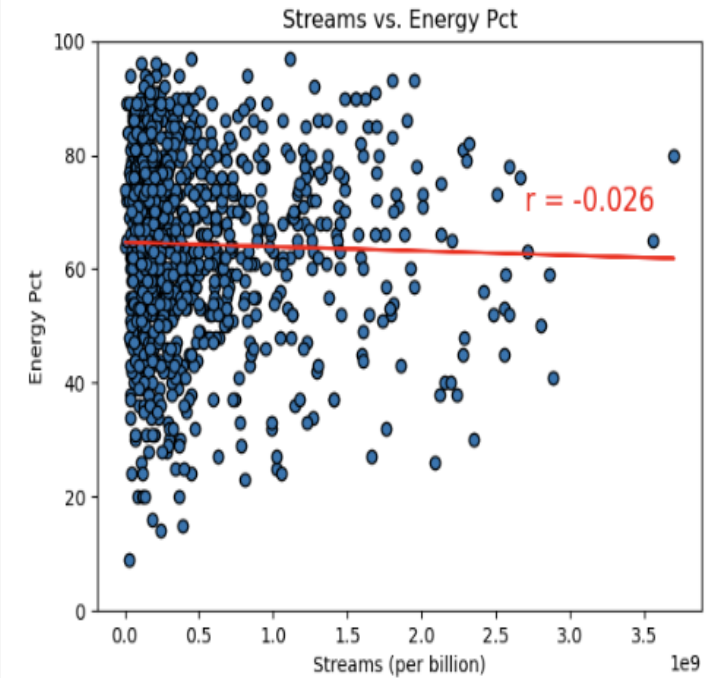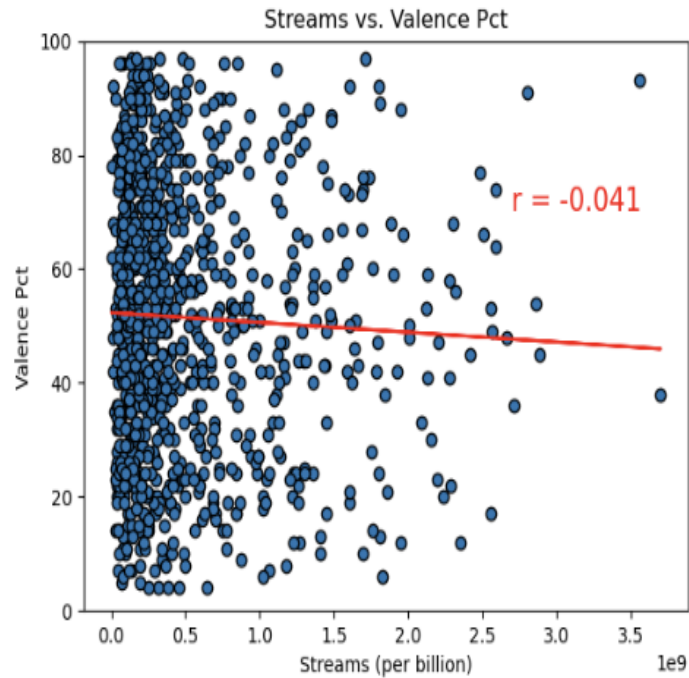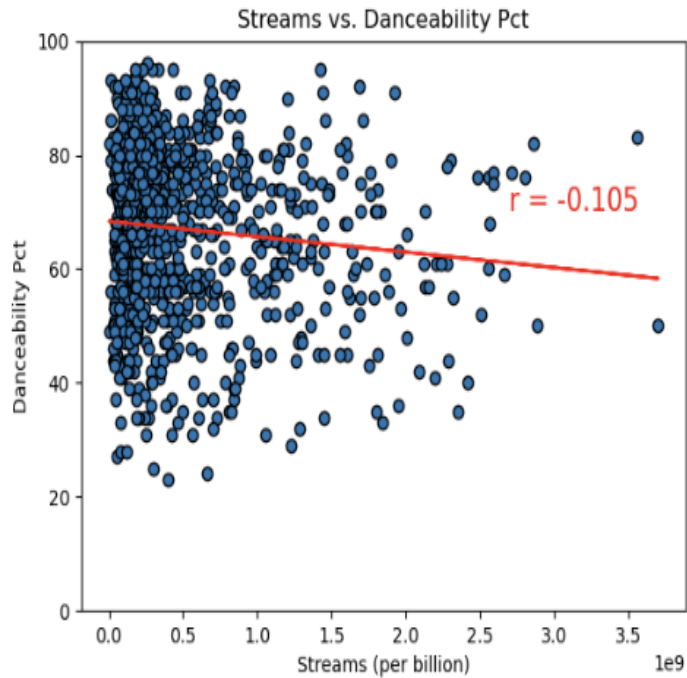
FINDING A PATTERN
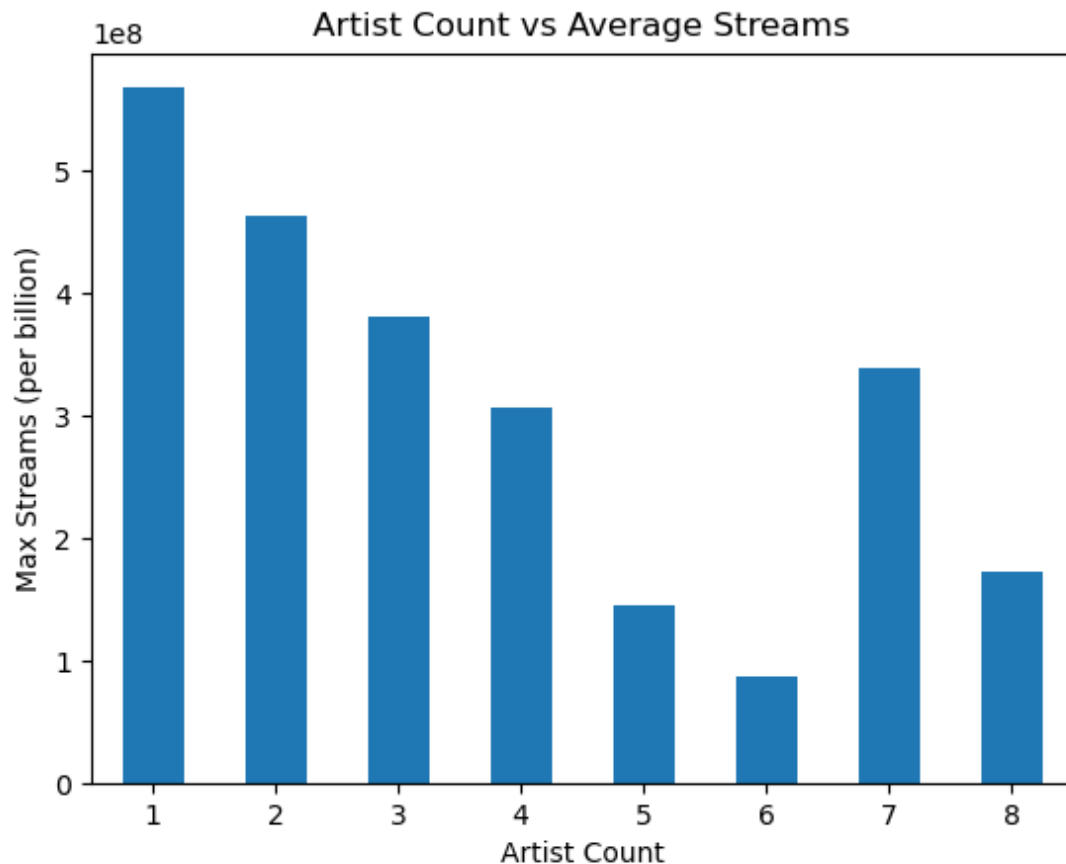
FINDING A PATTERN

FINDING A PATTERN

Amount of Songs Released per Month
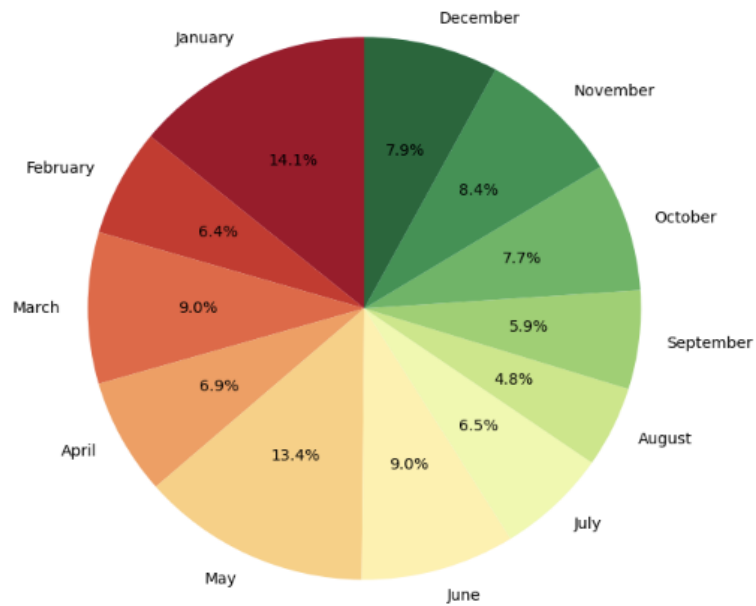
Amount of Streams per Month on Average

FINDING A PATTERN

Amount of Streams per Decade in Total
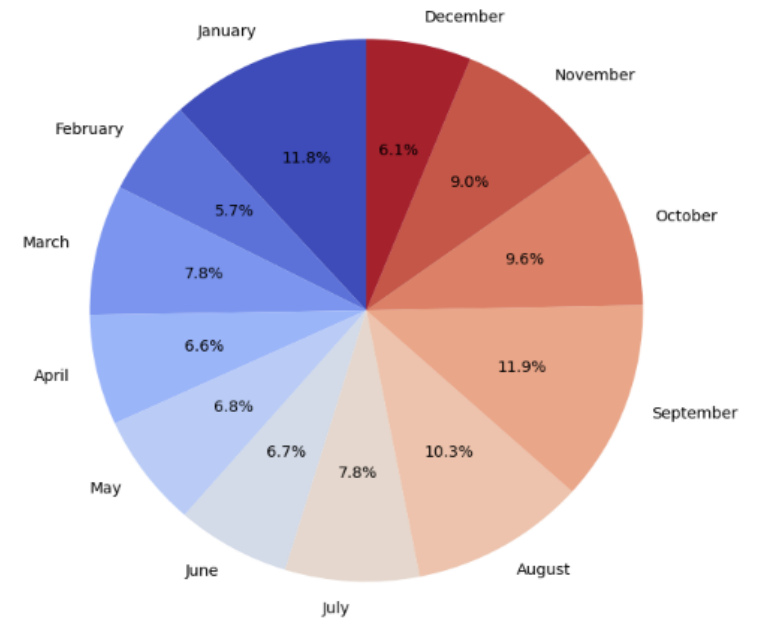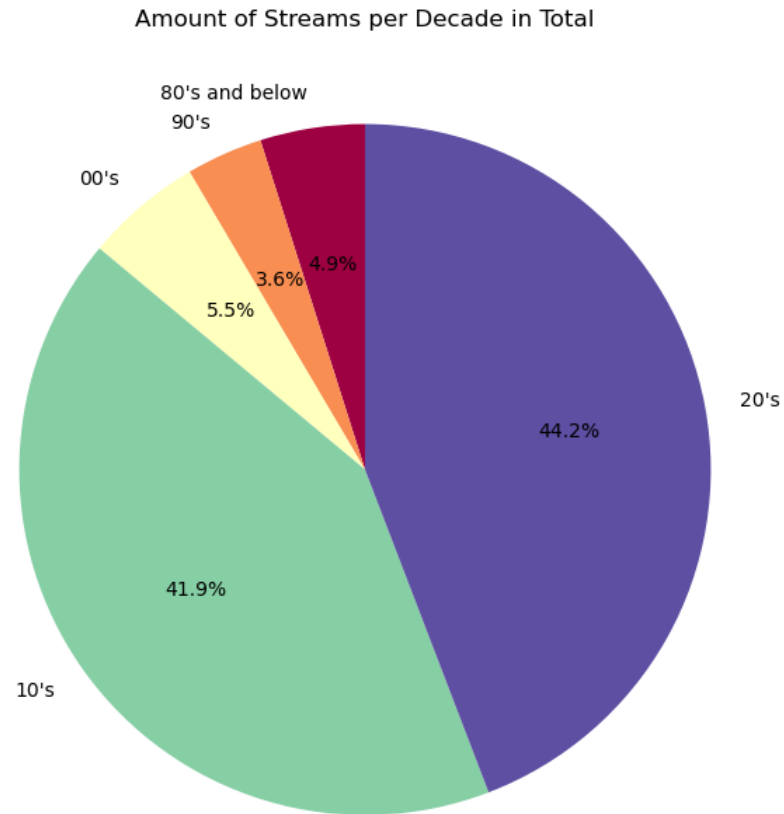
Decision Tree Model
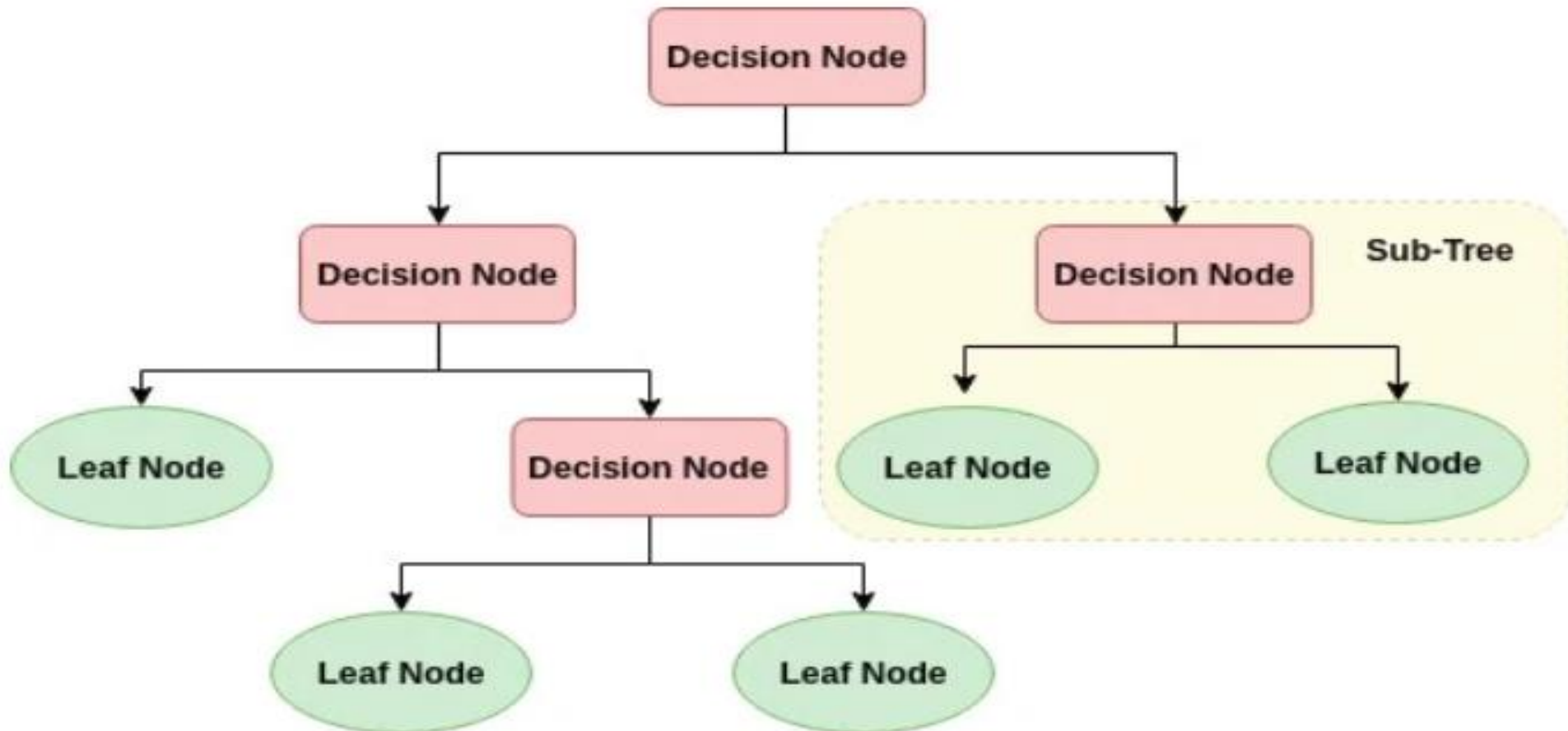
Decision Tree Model

```
Confusion Matrix

              Predicted 0    Predicted 1    Predicted 2

Actual 0          182             16              3

Actual 1           17             11              2

Actual 2            2              4              2
```

Decision Tree Model

```
Accuracy Score : 0.8158995815899581
Classification Report
              precision    recall  f1-score   support

           0       0.91      0.91      0.91       201
           1       0.35      0.37      0.36        30
           2       0.29      0.25      0.27         8

    accuracy                           0.82       239
   macro avg       0.52      0.51      0.51       239
weighted avg       0.82      0.82      0.82       239
```

Decision Tree Model Optimization

```
stream_category
0    201
2    201
1    201
Name: count, dtype: int64
```

Confusion Matrix

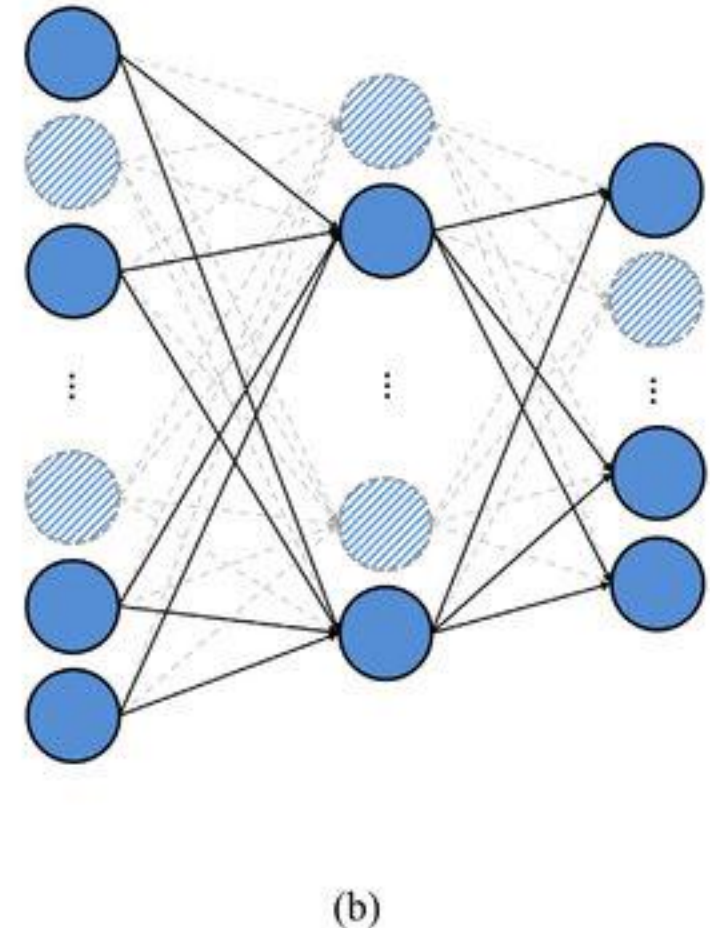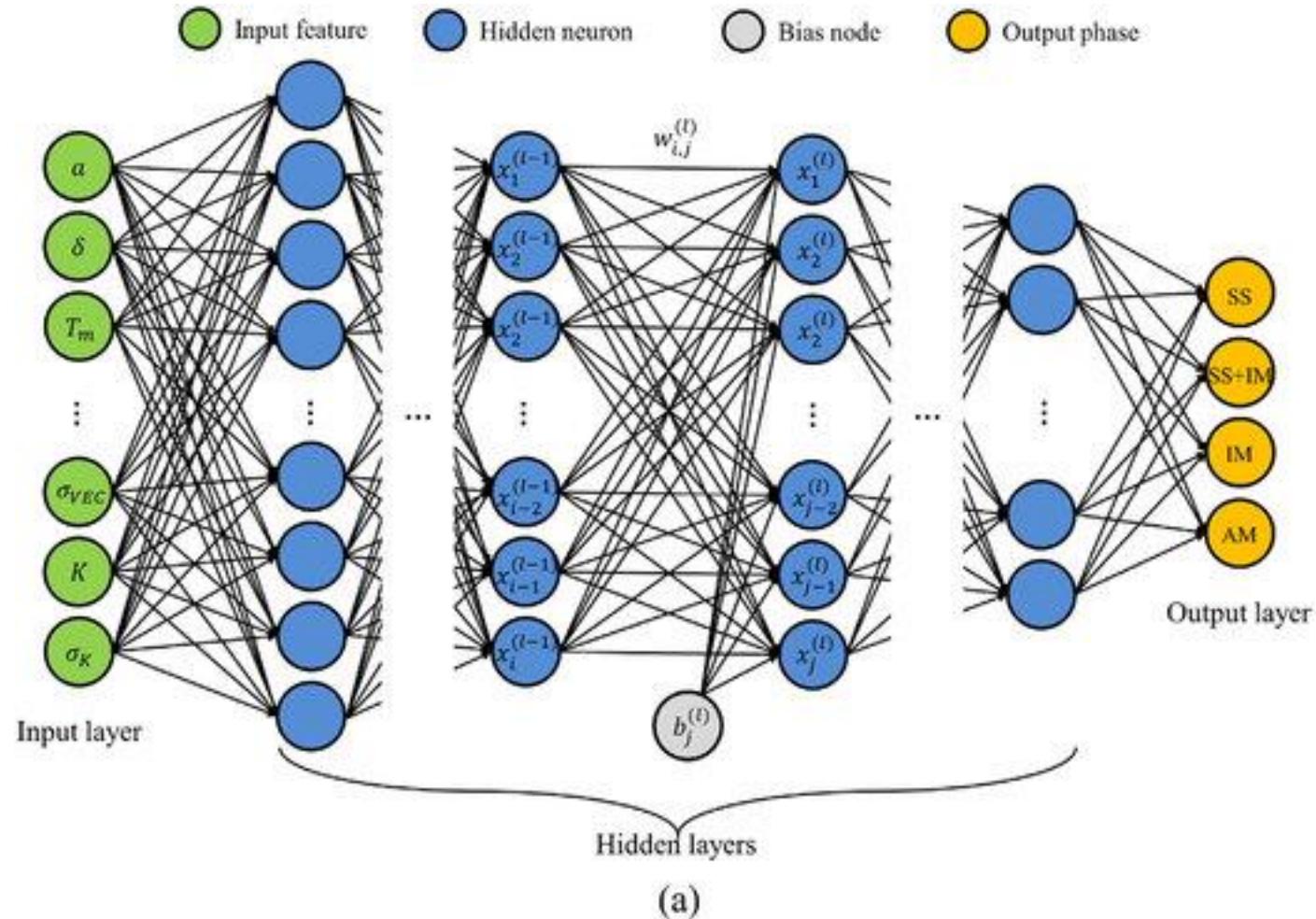|	Predicted 0	Predicted 1	Predicted 2
Actual 0	127	36	38
Actual 1	5	18	7
Actual 2	2	0	6

Accuracy Score : 0.6317991631799164
Classification Report

	precision	recall	f1-score	support
0	0.95	0.63	0.76	201
1	0.33	0.60	0.43	30
2	0.12	0.75	0.20	8
accuracy			0.63	239
macro avg	0.47	0.66	0.46	239
weighted avg	0.84	0.63	0.70	239

Neural Network Model

Neural Network Model

```
Model: "sequential"

_____
 Layer (type)                 Output Shape              Param #
=================================================================
 dense (Dense)                (None, 50)                750

 dense_1 (Dense)              (None, 15)                765

 dense_2 (Dense)              (None, 1)                 16


=================================================================
Total params: 1531 (5.98 KB)
Trainable params: 1531 (5.98 KB)
Non-trainable params: 0 (0.00 Byte)
_____
```

```
8/8 - 0s - loss: 0.6350 - accuracy: 0.7113 - 175ms/epoch - 22ms/step
Loss: 0.63502490504773, Accuracy: 0.7112970948219299
```

THANK YOU!