QUERIES FOR Segmentation & ANALYSIS IN SQL

1. **Segmentation :-**

WITH base AS (

    SELECT

        \*,

        CASE

          WHEN views IS NULL OR views = 0 THEN 0

          ELSE (likes + comments) / views

        END AS engagement\_ratio

    FROM `first-447405.Spotify\_Music\_Analysis.Music Data`

),

ranked AS (

    SELECT

        \*,

        PERCENT\_RANK() OVER (ORDER BY stream DESC) AS stream\_percentile,

        PERCENT\_RANK() OVER (ORDER BY engagement\_ratio DESC) AS engagement\_percentile

    FROM base

)

SELECT

    \*,

    CASE

        WHEN stream\_percentile <= 0.2 AND engagement\_percentile <= 0.2

            THEN 'Hit'

        WHEN stream\_percentile <= 0.2 AND engagement\_percentile >= 0.8

            THEN 'Shallow (Hype but low engagement)'

        WHEN stream\_percentile >= 0.8 AND engagement\_percentile <= 0.2

            THEN 'Hidden Gem'

        WHEN stream\_percentile >= 0.8 AND engagement\_percentile >= 0.8

            THEN 'Flop'

        ELSE 'Normal / Mid'

    END AS segment

FROM ranked

ORDER BY segment, stream DESC, engagement\_ratio DESC;

 -------------------------------------------------------------------------------------------------------------------------------------

**II)Hypothesis Testing:-**

1. **Does having a licence affect popularity of a song?**

SELECT Licensed,

avg(Stream) as average\_stream,

avg(engagement\_ratio) as engagement\_ratio

 FROM `first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data`

 GROUP BY 1

 -------------------------------------------------------------------------------------------------------------------------------------

1. **Does having a video affect popularity of a song**

SELECT official\_video,

avg(Stream) as average\_stream,

avg(engagement\_ratio) as engagement\_ratio

 FROM `first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data`

 GROUP BY 1

--------------------------------------------------------------------------------------------------------------------------------------

1. **Does platform affect popularity of a song?**

SELECT most\_playedon, avg(Stream) as average\_stream,

avg(engagement\_ratio) as engagement\_ratio,

count(\*) as count\_per

 FROM `first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data`

 GROUP BY 1

-------------------------------------------------------------------------------------------------------------------------------------

1. **Whether Duration of songs has any affect on it beign a hit or not**

SELECT

    CASE

        WHEN duration\_min <= 3 THEN 'Short'

        WHEN duration\_min > 3 AND duration\_min <= 5 THEN 'Medium'

        ELSE 'Long'

    END AS Length\_of\_song,

    segment,

    COUNT(\*) AS song\_count

FROM `first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data`

GROUP BY Length\_of\_song, segment

ORDER BY Length\_of\_song

-------------------------------------------------------------------------------------------------------------------------------------

1. **Do 'hits' sustain engagement longer than 'shallow'**

SELECT segment,

avg(danceability) as danceability,

avg(energy) as energy,

avg(loudness) as loudness,

avg(speechiness) as speechiness,

avg(acousticness) as acousticness,

avg(liveness) as liveness,

avg(valence) as valence,

avg(tempo) as tempo,

avg(instrumentalness) instrumentalness,

avg(EnergyLiveness) as EnergyLiveness

FROM first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data

GROUP BY 1

-------------------------------------------------------------------------------------------------------------------------------------

1. **Artist concentration i.e. artist with atleast more than 3 songs**

SELECT

  artist,

  COUNT(\*) AS unique\_songs,

  AVG(stream) AS avg\_stream,

  AVG(engagement\_ratio) AS avg\_engagement

FROM `first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data`

GROUP BY artist

HAVING COUNT(DISTINCT track) >= 3

ORDER BY 2 ;

-------------------------------------------------------------------------------------------------------------------------------------

1. **Engagment vs stream**

SELECT corr(engagement\_ratio,stream)

FROM `first-447405.Spotify\_Music\_Analysis.Segmented\_Music\_Data`

-------------------------------------------------------------------------------------------------------------------------------------