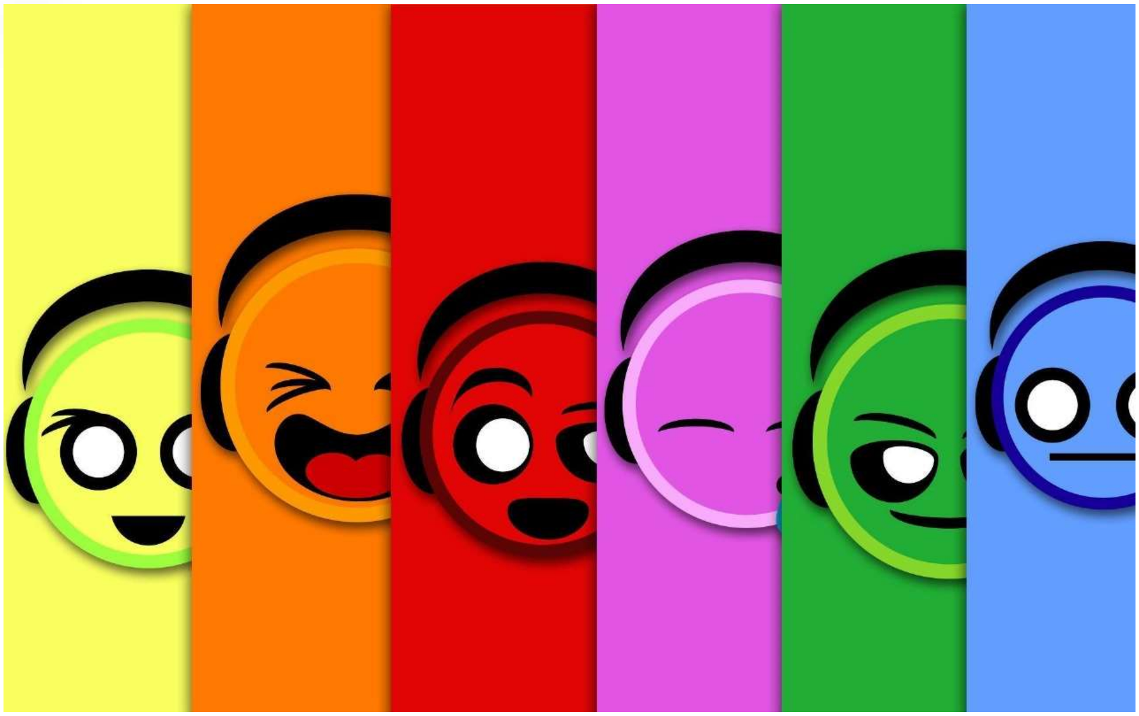
Does the Music You Love Impact Your Mental Health? A Data Analysis Perspective

Samarth Patel



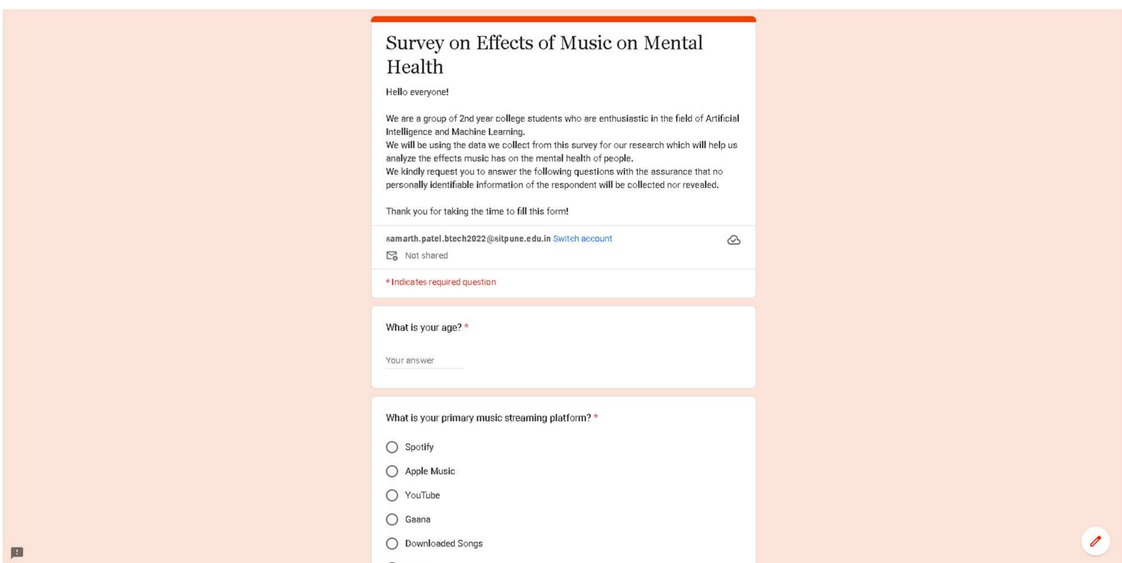
Think about your favourite song—the one that never fails to lift your spirits or bring a sense of calm. It's not just about catchy tunes; it seems to have a direct line to our emotions and mental state. This exploration is all about that connection between what we listen to and how it might affect our mental health. Using data-driven approach we're diving into the numbers to uncover if our music choices have a say in our psychological well-being. It's like detective work, but with statistics and playlists. The goal? To tease out any clues or patterns that reveal just how deeply music intertwines with our mental wellness and day to day behaviour.

How does music shape our mood?

Music is like a background pass to our emotions and mental well-being. It has a magical effect of bringing a sense of peace or get us grooving. After a stressful day, that one song might be the thing that brings your mind to peace. But why does music have such an impact on our health? Well, its more than just the beats and tempo of the song. Research shows that certain melodies, rhythms and lyrics can stimulate our brain, reducing stress and even enhancing our focus. In our research we try to link people’s music choices and habits with their day-to-day mental health. We try to form a link between how a person’s mood is based on the type of music they listen to.

Dataset Formation & Preprocessing

In our exploration of the ties between music and mental well-being, we decided to go with something user-friendly and accessible: Google Forms. We created this neat questionnaire that delved into people's musical tastes and how that might sway their mental game. People generously shared their top music genres, hangout spots for tunes, and the languages that hit them just right, but we didn't stop there.



Google Form to collect data

We wanted to see how different tunes make them feel – whether it jazzes them up or brings on the mental peace. And to dive deeper, we checked in on the mental side of things – like how often they face moments of panic, worry, lingering blues, fatigue, and losing interest. These responses are like gold dust. They give us this amazing mix of experiences to sift through, helping us spot potential links between musical choices and mental well-being indicators.

Surprisingly, a few participants entered ages that seemed to defy time itself - with entries like "age 1000" making an appearance. Additionally, we encountered a variation in data input styles, where some contributors logged their music-listening duration in minutes while others opted for the more leisurely approach in hours. These small yet unexpected bumps shed light on the practical challenges of real-life data collection. They served as a valuable reminder that such diversions are part of the process, emphasizing the importance of flexibility and readiness for surprises in the world of data acquisition.

We further broadened our horizons by incorporating a pair of additional datasets. One dataset comprises information about health professionals, a comprehensive crowd-sourced compilation of non-judgmental mental health practitioners in India, meticulously organized by state and city. The covered regions include Delhi NCR, Mumbai, Chennai, Kolkata, Bangalore, Assam, Hyderabad, Punjab, Madhya Pradesh, Maharashtra, Meghalaya, Rajasthan, and Tamil Nadu. The other dataset? It's all about the latest trending songs for each genre in our lineup, directly sourced from Spotify. With mental health professionals on one side and a collection of today's top music tracks on the other, our analysis is geared up for a unique blend of insights!

	Timestamp	age	streaming	ours_spen	_while_w	vorite_gen	plore_gen	ther_genr	er_langua	nxiety_lev	ression_l	omnia_le	ocd_level	usic_effe	ermission
0	2023-09-02 20:21:54	18	Spotify	3	Yes	Pop	Sometime	Pop, Bolly	Punjabi	2	0	1	3	Improves	I understa
1	2023-09-02 20:22:14	19	Apple Mu	3	Yes	Hip-Hop/F	Sometime	Pop, Bolly	Punjabi	0	0	0	0	Improves	I understa
2	2023-09-02 20:23:07	18	Spotify	1	No	Pop	Sometime	Bollywooc	No other l	4	1	1	6	No effect	I understa
3	2023-09-02 20:23:18	19	Spotify	4	No	R&B (Rhyt	Yes	Pop, Bolly	Punjabi, K	7	5	6	3	Improves	I understa
4	2023-09-02 20:24:09	18	Spotify	2	No	indie	Sometime	Bollywooc	Spanish	1	0	0	0	Improves	I understa
5	2023-09-02 20:25:30	19	Spotify	0.5	No	Pop	Yes	Pop, Bolly	Punjabi, S	8	5	3	4	Improves	I understa
6	2023-09-02 20:25:33	21	Spotify	3	Yes	Bollywooc	Sometime	Hip-Hop/F	Punjabi	8	4	3	4	Improves	I understa
7	2023-09-02 20:25:35	17	Spotify	3.5	Yes	Pop	Yes	Bollywooc	Korean, Fr	8	5	0	1	Improves	I understa
8	2023-09-02 20:25:55	19	Spotify	2	No	Bollywooc	Yes	Classical	Punjabi	9	3	6	7	Improves	I understa
9	2023-09-02 20:25:55	18	Spotify	0.5	No	Bollywooc	Yes	Pop, Hip-H	Punjabi, T	1	0	0	0	Improves	I understa
10	2023-09-02 20:25:56	17	Spotify	4	Yes	Pop	Yes	Bollywooc	Punjabi, S	7	4	5	1	Improves	I understa
11	2023-09-02 20:26:12	20	YouTube	0.5	No	Electronic	Yes	Bollywooc	No other l	5	3	7	2	Improves	I understa
12	2023-09-02 20:26:27	18	Spotify	2	Yes	Bollywooc	Sometime	Rock, Clas	No other l	4	4	3	1	Improves	I understa
13	2023-09-02 20:27:02	19	Download	2	Yes	Rock	Sometime	Pop, Bolly	No other l	4	0	0	0	Improves	I understa
14	2023-09-02 20:27:04	19	Spotify	1	Yes	Electronic	Sometime	Pop, Rock	No other l	3	5	8	0	Improves	I understa
15	2023-09-02 20:27:53	19	Spotify	5	Yes	Bollywooc	Sometime	Bollywooc	Punjabi, T	4	2	1	1	Improves	I understa
16	2023-09-02 20:28:01	18	Spotify	1	No	Pop	Yes	Bollywooc	Korean, Fr	1	0	2	0	Improves	I understa
17	2023-09-02 20:29:35	18	Spotify	2	No	Pop	Sometime	Bollywooc	Spanish, K	3	0	1	2	Improves	I understa
18	2023-09-02 20:30:04	18	Spotify	3	Yes	Rock	Yes	Pop, Hip-H	No other l	7	7	0	3	Improves	I understa
19	2023-09-02 20:30:05	19	Spotify	3	Yes	Pop	Yes	Bollywooc	No other l	8	9	7	7	Improves	I understa
20	2023-09-02 20:31:09	52	Spotify	5	Yes	Electronic	Yes	Devotiona	Spanish, K	7	7	7	7	Improves	I understa
21	2023-09-02 20:34:01	19	Spotify	1	Yes	Rock	Sometime	Bollywooc	Telegu	7	4	4	4	Improves	I understa
22	2023-09-02 20:34:37	42	Spotify	2	Yes	Bollywooc	Yes	Pop, Hip-H	Punjabi	5	5	8	3	Improves	I understa
23	2023-09-02 20:34:44	17	Spotify	1.5	Yes	Rock	Yes	Pop, Bolly	Punjabi, K	8	7	2	7	Improves	I understa
24	2023-09-02 20:35:20	18	Spotify	3	Yes	Pop	Yes	Bollywooc	No other l	3	0	0	0	Improves	I understa
25	2023-09-02 20:35:51	54	YouTube	3	Yes	Bollywooc	Yes	Bollywooc	Punjabi	1	1	1	0	Improves	I understa

Primary Dataset

Moving onto the Data preprocessing phase, our first task involved revamping our dataset by changing the column names. This is done to enhance the readability of the dataset. The original column names were lengthy, overly technical and might cause confusion during data analysis.

Next, it was time to tackle those troll entries - like when someone entered an age of 1000 and such. Considering the oldest person on record lived to be 120 years old (and age can't be negative), we had to bid farewell to those rows that didn't quite match the mark.

```
age_criteria = df[ (df['age']>120) | (df['age']<0) ].index
df.drop(age_criteria, inplace=True)
```

Code to handle troll value related to age

Additionally, we had to tackle the 'hours_spent' field. We presumed that entries like 30, 40, or 45 were unlikely to represent hours, so we made necessary adjustments to ensure accuracy in the dataset.

```
def min_hr(mins):
  hr = mins/60
  return hr

for i in df.index:
  hours = df.loc[i,'hours_spent']
  if hours >= 10:
    if (hours%5==0):
      df.loc[i,'hours_spent'] = min_hr(hours)
    else:
      df.drop(i, inplace=True)
```

Code to handle hours_spent field

Next, lets zoom into one particular column, ‘other_genre’. In our Google Form we asked people to list all the genres they listen to except their favourite one (we had that covered separately). There were heaps of options to pick from, plus a free space to write down any other genre that wasn’t already included into above options. Here’s the deal: Some of the entries in the provided free space were basically the right genres but spelled incorrectly, while others were subgenres that were already in the options. So now we had to map these entries to already existing ones or create new ones.

```
for genre in genre_map:
  for i in df.index:
    if df.loc[i,'favorite_genre'] in genre_map[genre]:
      df.loc[i,'favorite_genre'] = genre
```

Mapping genres in column other_genre

Moving on, we faced a similar issue with ‘primary__streaming__platform’ that we encountered in ‘other__genre’. So we had to map those entries to existing or create new entries accordingly.

```
for sp in streaming_platform_mapping:
  for i in df.index:
    if df.loc[i,'primary_streaming_platform'] in
streaming_platform_mapping[sp]:
      df.loc[i,'primary_streaming_platform'] = sp
```

Mapping streaming platforms in column primary__streaming__platform

Moving on, given we provided users the options to select multiple options in ‘other__genre’ and ‘other__languages’ we created in individual columns for each genre and language. This here is to access how many individuals were drawn to each genre or language.

```
genres_list = ['Pop', 'Bollywood', 'Hip-
Hop/Rap','Rock','Devotional','Classical','Instrumental',
               'R&B (Rhythm and Blues)','Folk', 'Electronic/Dance','No
other genres']

for genre in genres_list:
  for i in df.index:
    if genre in df.loc[i,'other_genres']:
      df.loc[i,genre] = 1
    else:
      df.loc[i,genre] = 0
```

Separating other__genre into individual columns

```
langs_list = ['Punjabi','Spanish','Korean','Tamil','Telegu',
              'Malayalam','French','Portugese','Chinese','No other
language']

for lang in langs_list:
  for i in df.index:
    if lang in df.loc[i,'other_languages']:
      df.loc[i,lang] = 1
    else:
      df.loc[i,lang] = 0
```

Separating other__languages into individual columns

Now the issue is some people had selected their favourite genre again in the 'other genres' question which could lead to redundancy in data.

So, we removed the value from the genre columns of a sample if it is already selected as the favourite genre. And if after that there is no other genre they listen to, we will add them to the 'no other genres' category.

```
for i in df.index:
  fav = df.loc[i,'favorite_genre']
  other = df.loc[i,'other_genres']
  if fav in other:
    df.loc[i,fav] = 0
  if any(df.loc[i,genres_list[:-1]])!=True:
    df.loc[i,genres_list[-1]] = 1
```

Handling favorite_genre to prevent data redundancy

And lastly, since anxiety and depression levels were self-reported, we could change them to more common everyday used terms related to the most common symptoms of these issues and also change them from integer score to categorized levels.

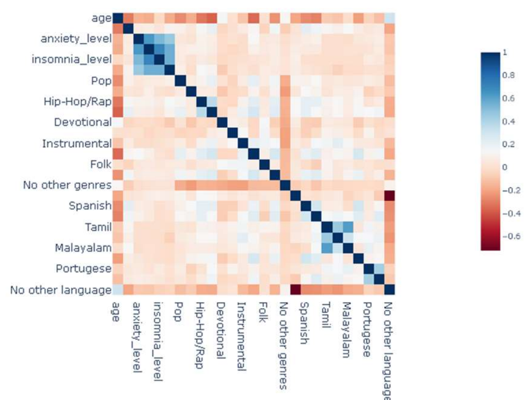
```
for i in df.index:
  level = df.loc[i,'anxiety_level']
  new = 'persistent_state_of_worry_panic_fear'
  if level == 0:
    df.loc[i,new] = 'none'
  elif level >= 1 and level <= 3:
    df.loc[i,new] = 'mild'
  elif level >= 4 and level <= 6:
    df.loc[i,new] = 'moderate'
  elif level >= 7 and level <= 9:
    df.loc[i,new] = 'severe'
  else:
    df.loc[i,new] = 'extreme'

for i in df.index:
  level = df.loc[i,'depression_level']
  new = 'persistent_sadness_tiredness_loss_of_interest'
  if level >= 0 and level <= 1.6:
    df.loc[i,new] = 'normal'
  elif level > 1.6 and level <= 2.4:
    df.loc[i,new] = 'mild'
```

```
  elif level > 2.4 and level <= 2.9:
    df.loc[i,new] = 'borderline'
  elif level > 2.9 and level <= 4.4:
    df.loc[i,new] = 'moderate'
  elif level > 4.4 and level <= 6.0:
    df.loc[i,new] = 'severe'
  else:
    df.loc[i,new] = 'extreme'
```

Integer scores to categorized levels for columns-anxiety and depression

Exploratory Data Analysis

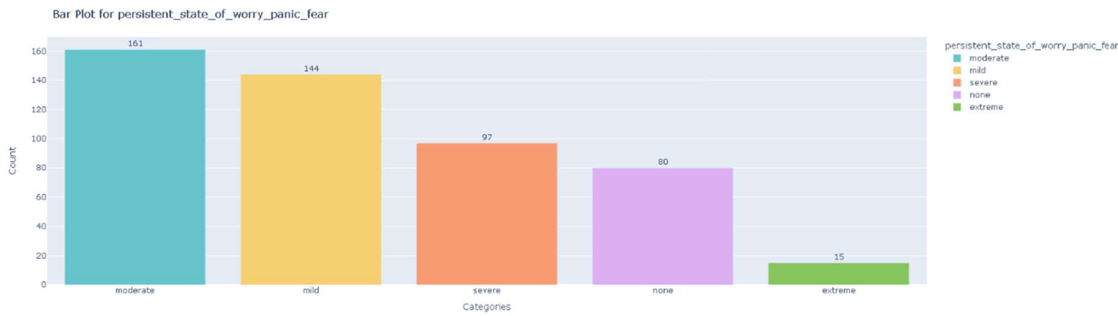
Moving onto our Exploratory data analysis. EDA is like peering into the treasure trove of information. In this we will be sifting through numbers and patters to uncover the fascinating stories and information our dataset holds.



Graph 1

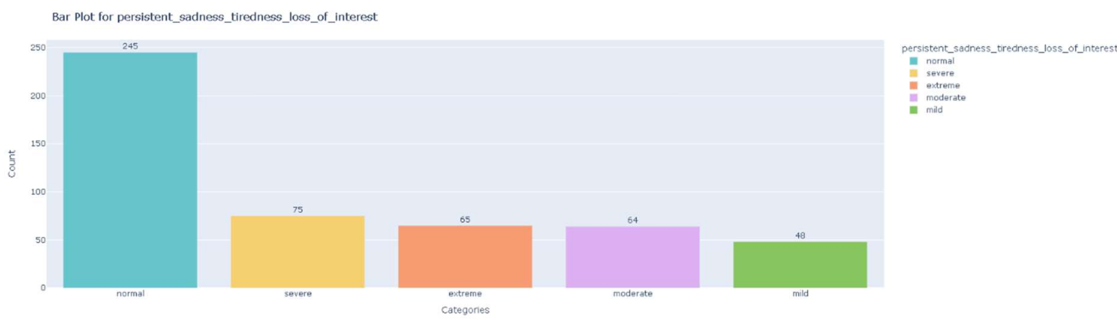
The first graph itself revels some intriguing connections within our data , for example , more than 50% if Tamil music listeners also tune into Telegu and Malayalam songs. This pattern mirrors with listeners of Portuguese and French as well.

Another interesting stat we found is genres like Hip-hop/Pop, rock and R&B have a negative correlation with age, meaning mostly younger people listen to these genre



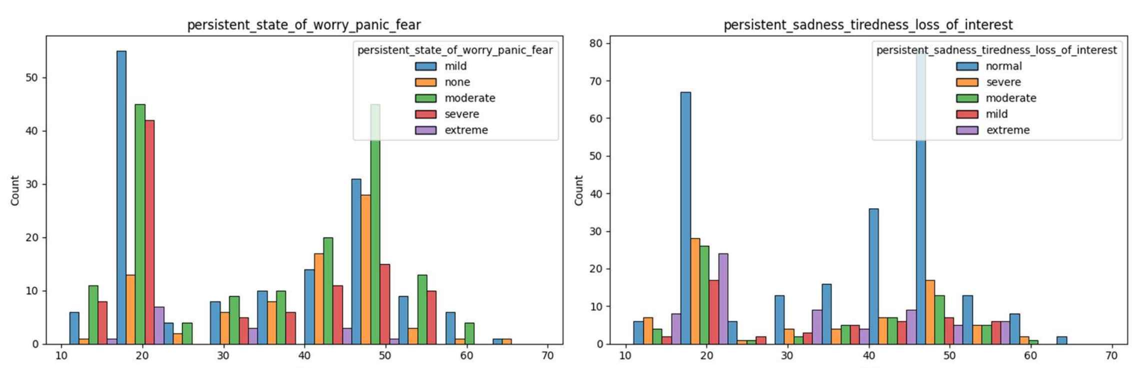
Graph 2

From graph 2, it seems like a good chunk of responders, about 32.39% are dealing with moderate worry/panic/fear, followed closely by 28.97% who reported mild levels for these emotions. On the flip side, extreme state of worry/panic/fear seems to be relatively low, just 15 out of 497 entries.



Graph 3

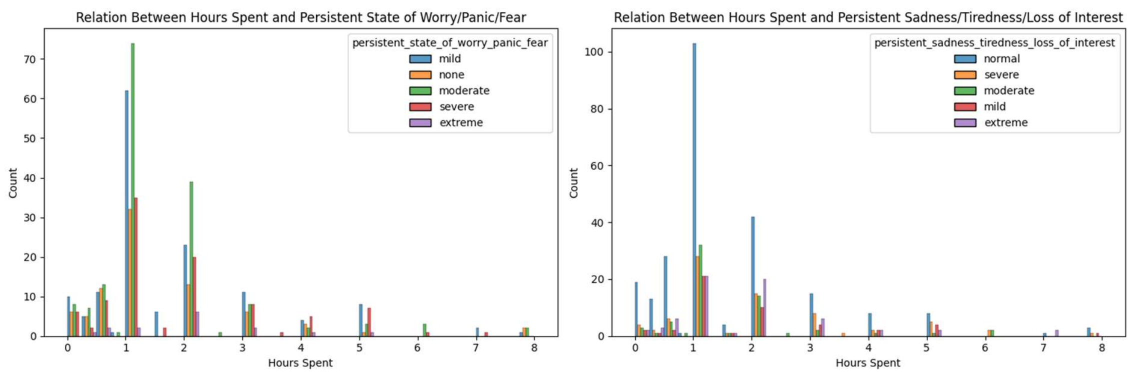
For Graph 3, majority of responders, about 49.29% of survey responders show what could be considered normal levels of sadness/tiredness/loss of interest. On the extreme side, about 13.07% of responders appear to be experiencing these feelings at an extreme level.



Graph 4

From graph 4 we could infer that most people of age group 15-17 show mild or moderate levels of worry/panic/fear levels. Whereas people of the age group 20-23 show more instances of severe worry/panic/fear.

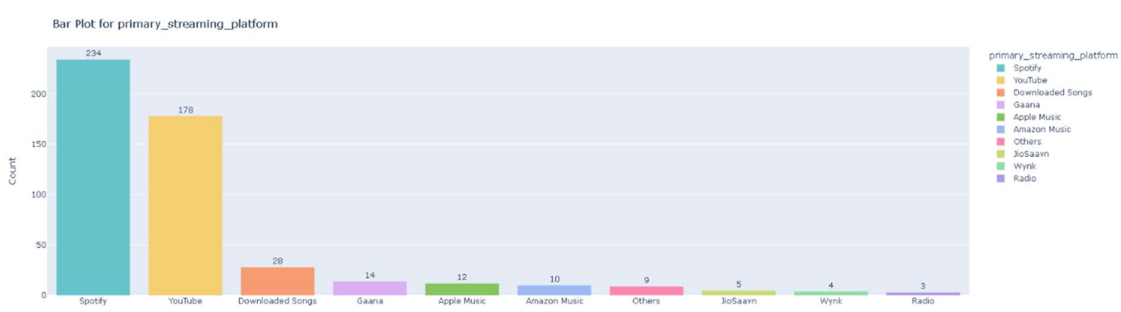
The people of the age group 15-17 & 45-50 show normal sadness/tiredness/loss of interest levels. Whereas people of the age group 20-23 show more instances of severe sadness/tiredness/loss of interest.



Graph 5

From Graph 5 we infer that most people who listen to music for around 1 hour show mild or moderate worry/panic/fear levels. Whereas people who listen to music for around 5 hours show more instances of severe worry/panic/fear.

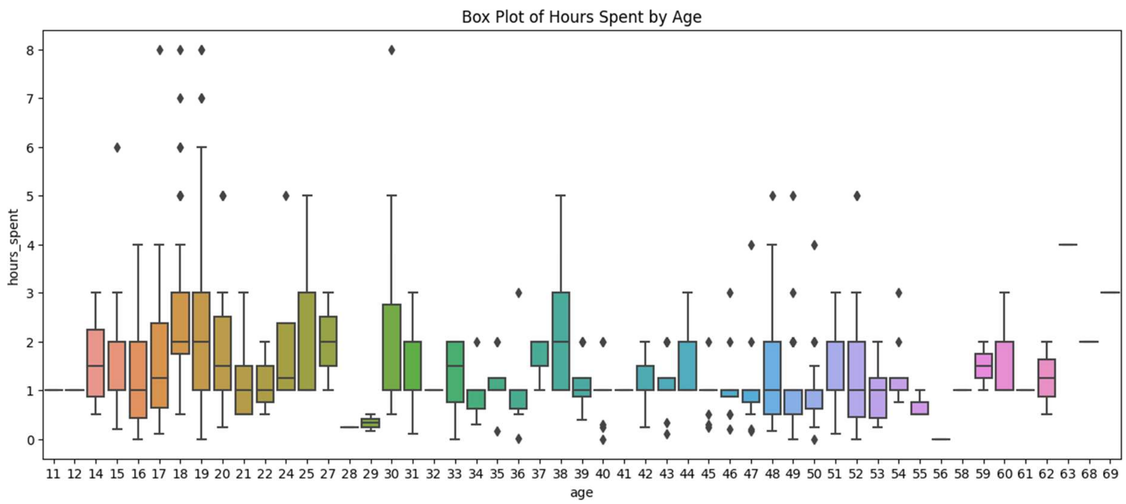
The people who listen to music for around 1 hour show normal sadness/tiredness/loss of interest levels. Whereas people who listen to music for around 5 hours show more instances of severe sadness/tiredness/loss of interest.



Graph 6

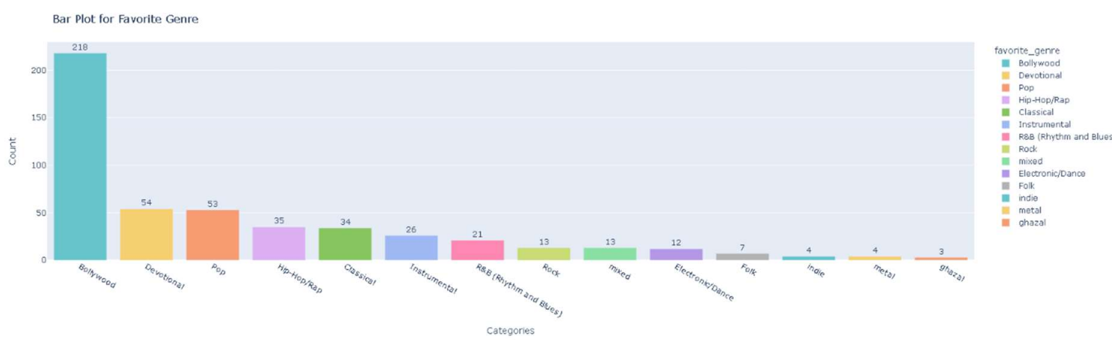
This plot helps us recognize that most of the people (234 out of 497, 47.08%) people use Spotify as their primary music streaming platform.

This position is followed by Youtube (178 out of 497, 35.81% of people).



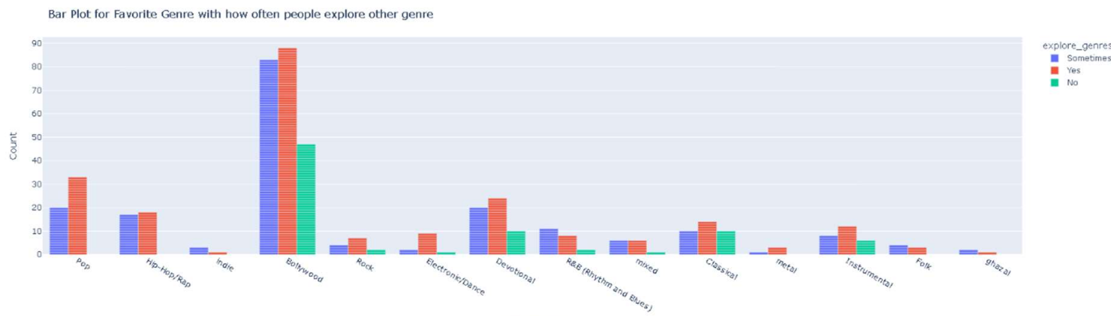
Graph 7

From Graph 7 From this plot, we understand the distribution and pattern of hours spent by people of various age groups. We can identify the outliers for each age group. The places we see only a dash signifies only one input for that age.



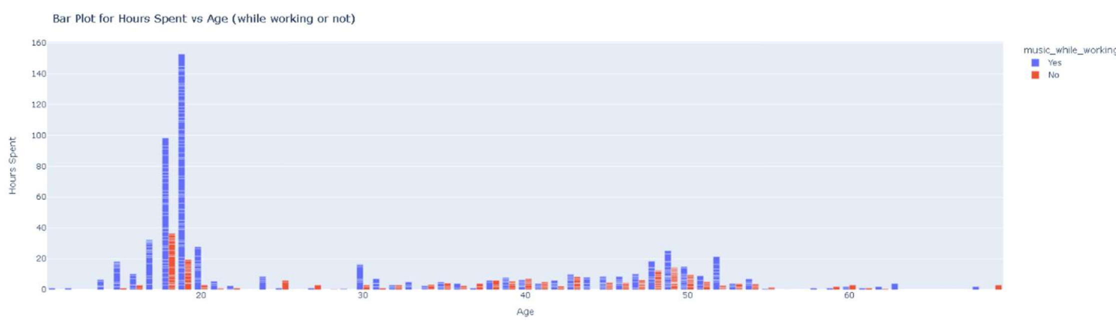
Graph 9

Graph 9 is a simple Bar Plot for Favourite Genre.



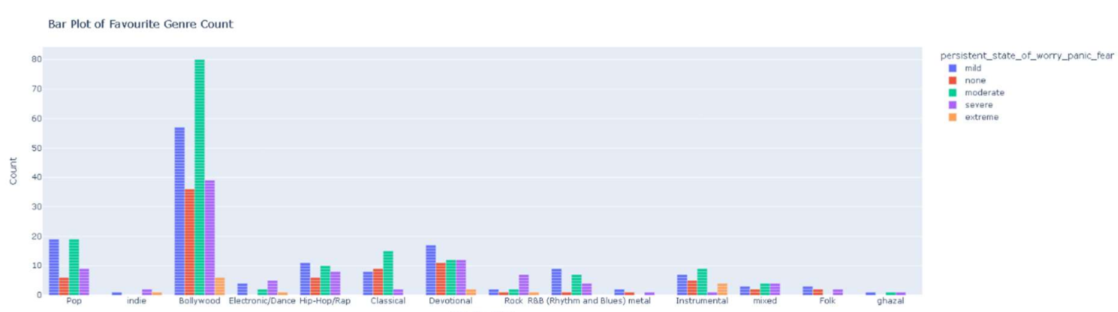
Graph 10

From this plot we can infer how often people listen and explore genres other than their favourite ones. Here we see that although Bollywood has the maximum fanbase, it also has most of its audience exploring other genres frequently. Yet Bollywood remains the favourite.



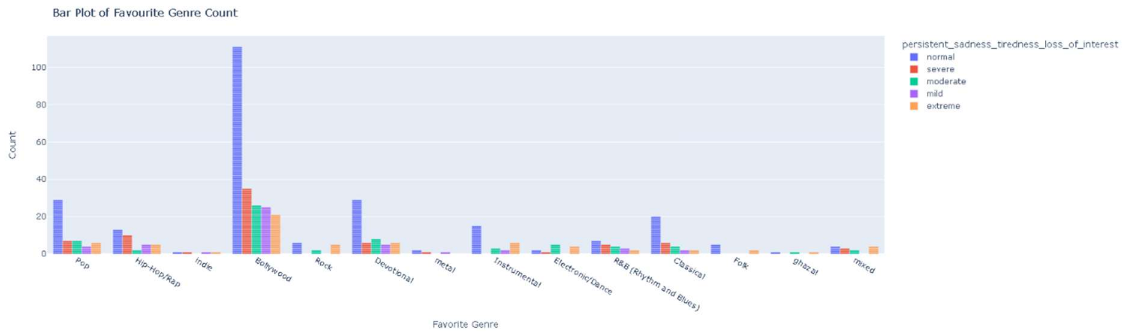
Graph 11

We infer from the above plot that people of the age group 15-20 are more likely to listen to music while working as compared to other people, hence their hours of listening to music is more as well.



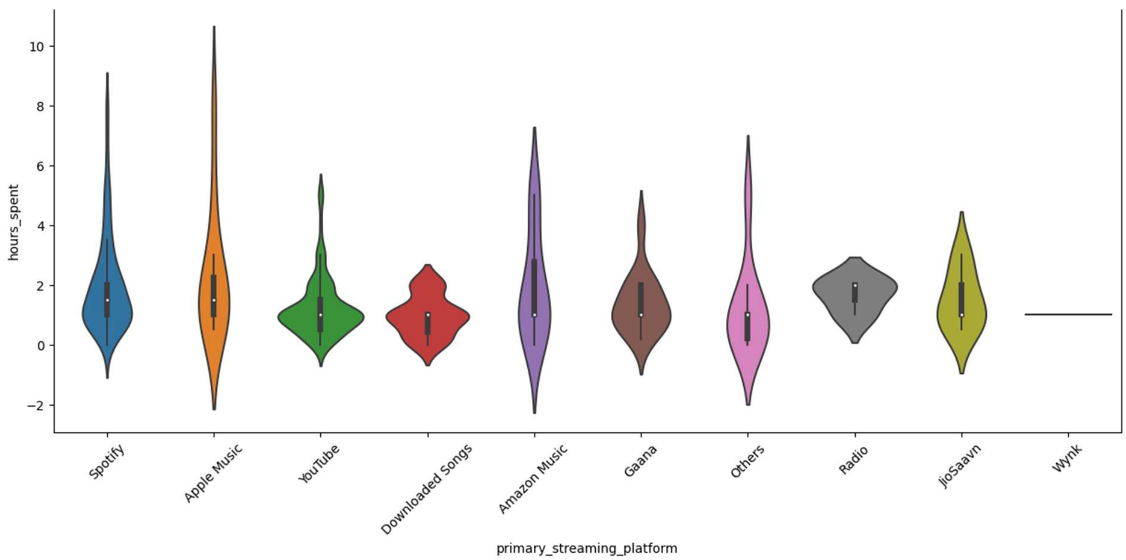
Graph 12

Graph 12 shows that majority of people with Moderate level of worry/panic/fear listen to Bollywood genre.



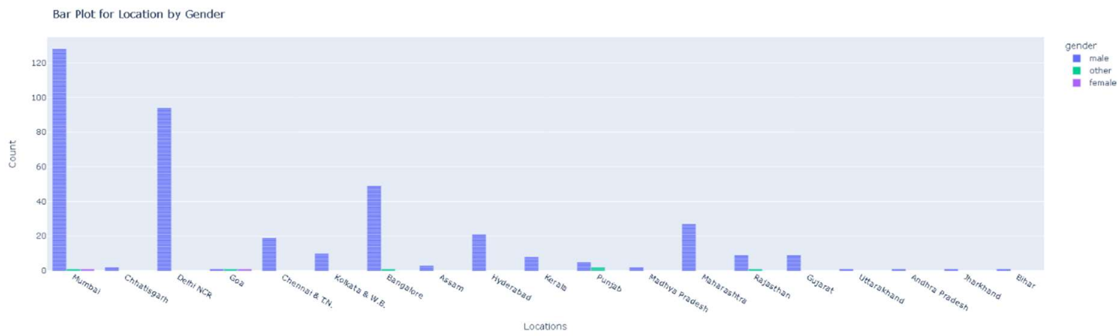
Graph 13

Graph 13 shows that majority of people with normal level of sadness/tiredness/loss of interest listen to Bollywood genre.



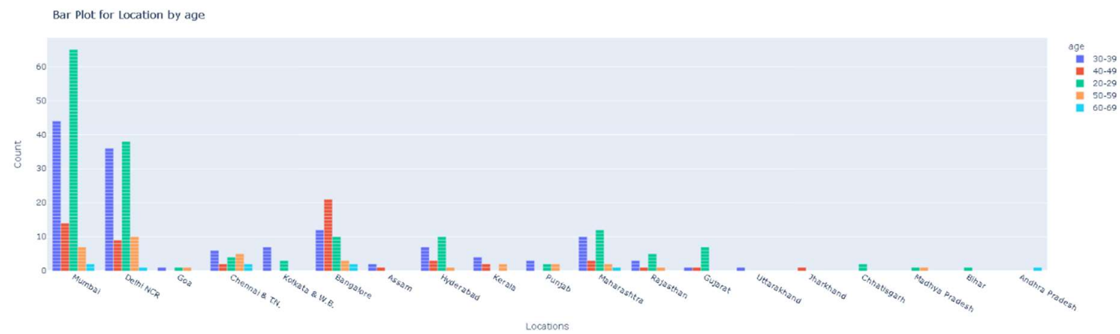
Graph 14

Graph 14 is a violin plot of Primary Streaming Platform according to Hours spent.



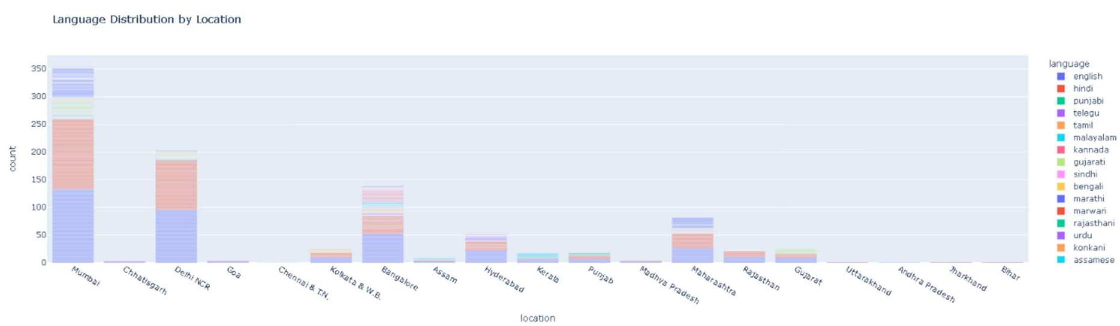
Graph 15

Graph 15 was performed on the Health Professionals Dataset. This particular plot shows that there are close to none female health professionals in our country.



Graph 16

In Graph 16 we can see that most of the mental health professionals in each location belong to the age group 20-29. Thus, showing the generation gap in mental health awareness and acceptance.



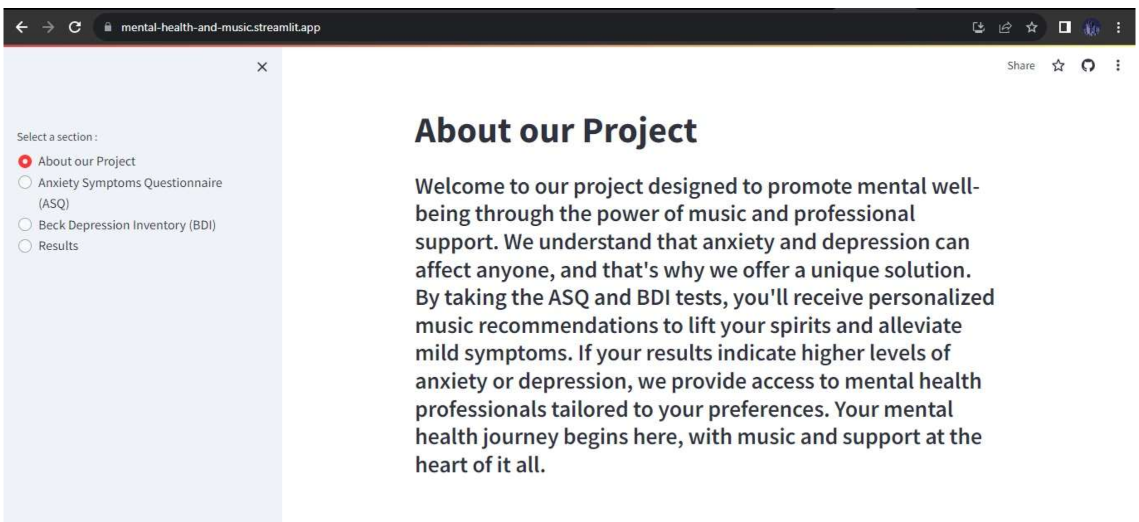
Graph 17

Graph 17 shows that the most frequent language spoken by health professionals in English followed by Hindi and the Regional Language.Fun fact the 4th most spoken language in Mumbai is Gujarati.

Deployment

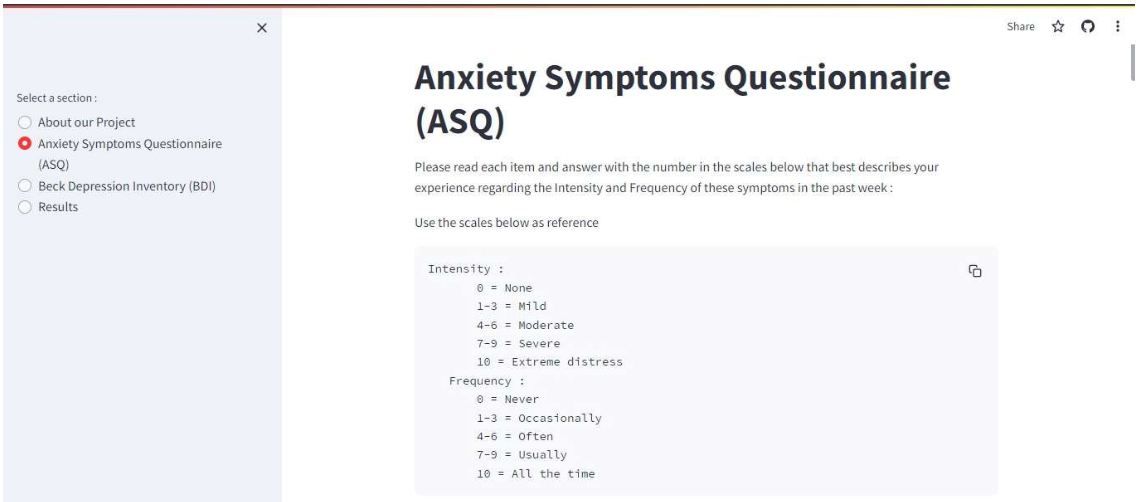
Finally in our deployment, we delve into the practical part of our analysis. We've developed a user-friendly streamlit app, which servers as a comprehensive space for mental health assessments and enjoying music suggestions. Picture it as your personal mental health check-up paired with Spotify's top hits. Our app seamlessly intertwines mental health evaluations with delightful musical recommendations, offering a complete experience at your fingertips.

Our Streamlit app is divided into 4 sections, each having their own purpose. This initial page serves as a hub, providing an overview of our research topic and the purpose of our study. It offers users a brief but comprehensive insight into the essence of our work.



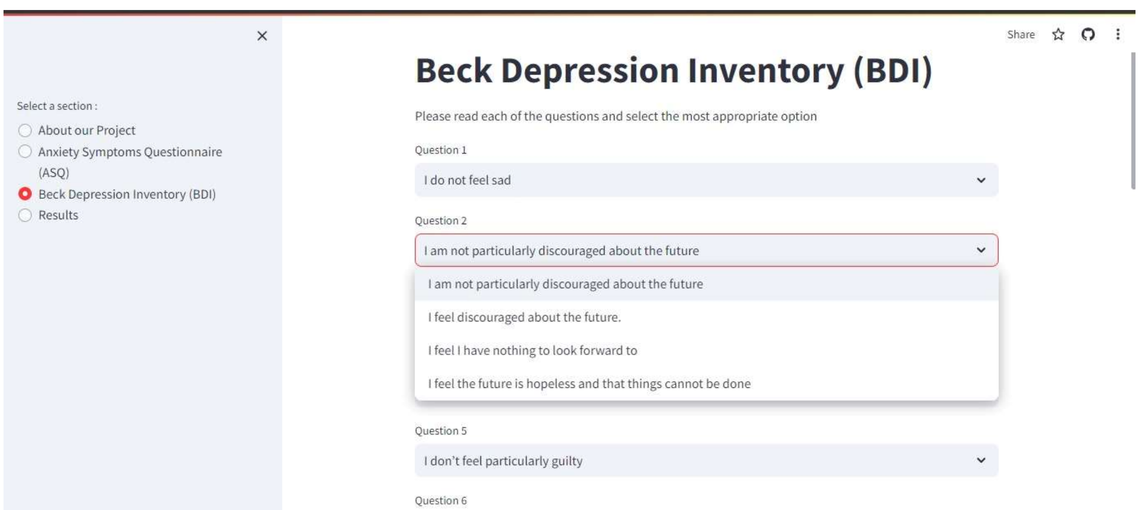
General Information Page

The second page is dedicated to the ASQ, a concise self-report questionnaire designed to evaluate the frequency and intensity of anxiety symptoms. Our implementation of the ASQ aims to improve the assessment of anxiety symptoms within clinical contexts, enhancing the understanding and measurement of anxiety.



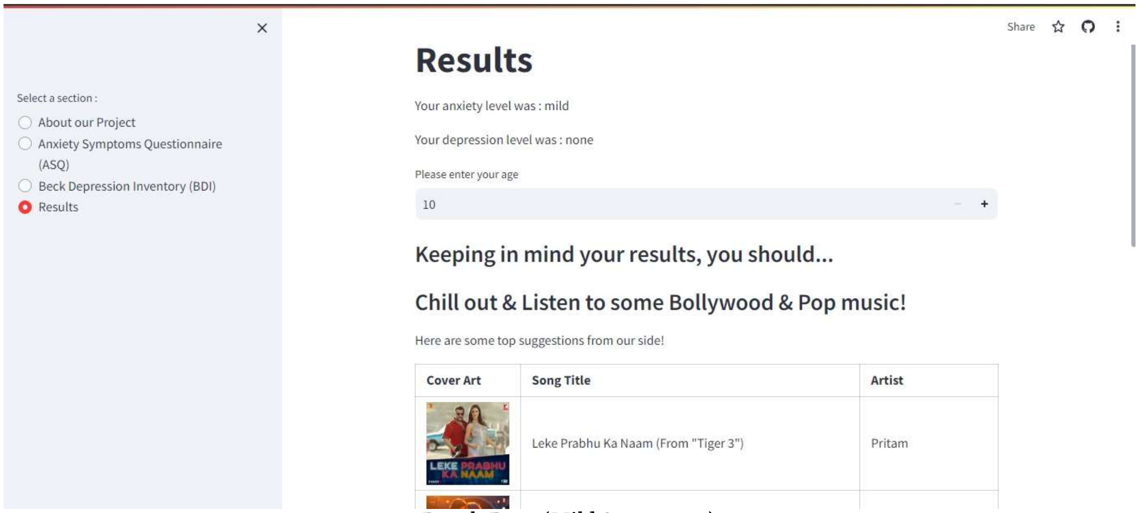
Anxiety Symptoms Questionnaire (ASQ) Page

The third page contains the Beck Depression Inventory, which is a 21-question multi-choice self-report inventory which is widely utilized in psychological assessments to gauge the severity of depression symptoms.

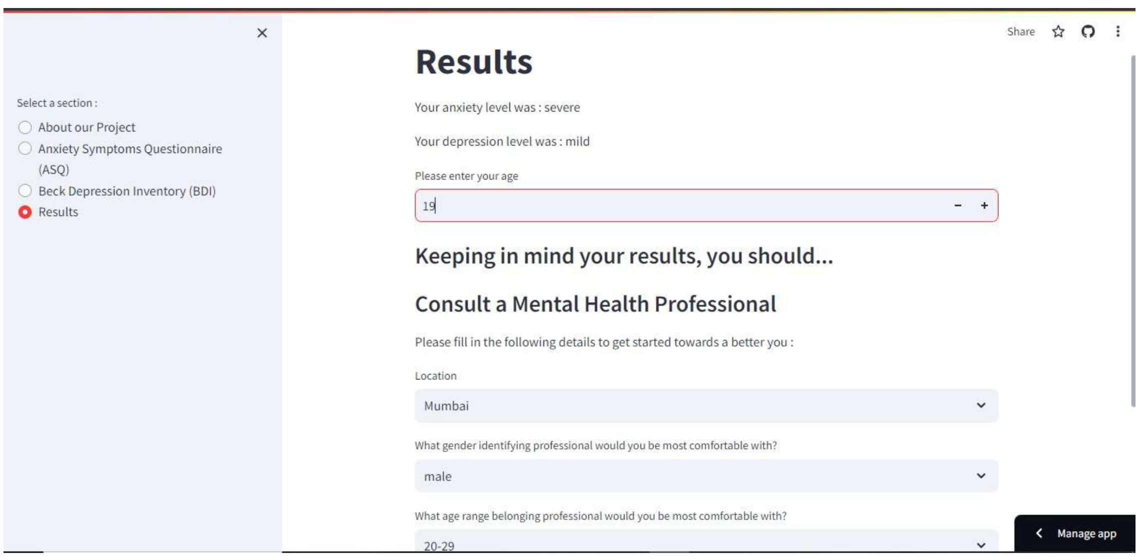


Beck Depression Inventory Page

The final page combines the outcomes from both the ASQ and Beck Depression Inventory. For users exhibiting severe symptoms, the app recommends nearby health professional whose details was stored in Health Professionals Dataset mentioned earlier. In contrast, for milder symptoms, the app steers users towards the music recommendations sourced from Spotify-trending playlists. This approach combines mental health assessment with a touch of music therapy catering to varying levels of needs all within a single interface.



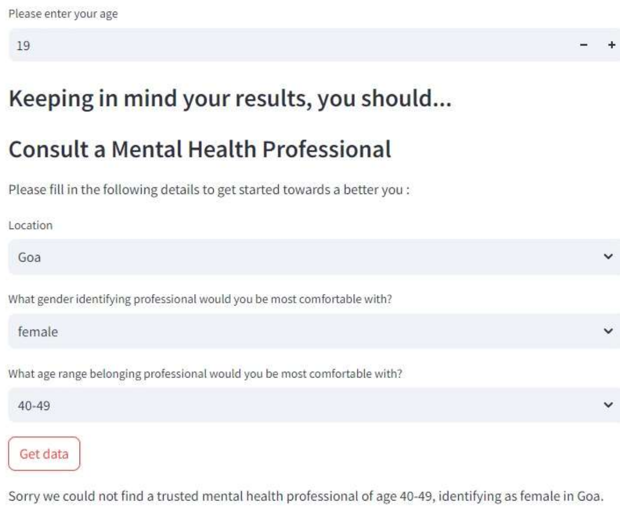
Result Page (Mild Symptoms)



Result Page (Severe Symptoms)

Drawbacks and Future Scope

So, talking about the not-so-great parts first – in our project, the health professionals' data has a broad stroke when it comes to locations. For instance, places like Maharashtra are covered as a whole, without city-specific details. It's like trying to find a needle in a haystack when you're looking for someone in a whole state. It also lacks diversity in its data.



Drawbacks Health Professionals Data

Another drawback is our primary dataset, it’s a bit small in size. More data would definitely be enable us to fine tune our assessments and help us get more in-depth look at mental health. The dataset also lacks diversity, for example majority age of people who filled our Google form fall below 30 years.

Looking ahead, one of our goals for this project is teaming up with health professionals and clinics. Imagine the power of combining forces! This collaboration could bring us a treasure trove of in-depth analysis, helping us zoom in even closer and boosting the accuracy in spotting mental health issues that people might be dealing with.

Conclusion

In conclusion, our journey exploring the relationship between music and mental health has been both enlightening and eye-opening. We've delved into the fascinating world where tunes and mental well-being intersect, uncovering valuable insights through data analysis. However, while our exploration has been insightful, it's just the tip of the iceberg in this vast field.

The connections we've unearthed between music preferences and mental health are intriguing, showing potential paths to understand and support mental well-being. The use of data-driven tools, like our assessment models and song recommendations, exemplifies the power of merging technology with mental health exploration