House_Cloud : Best DJ in the city

(SF electronic music events rating app)

Project Goal:

This work presents an app construction that is rating DJ artists of electronic music events in SF according to their number of followers on SoundCloud, then presents top clubs where DJs with the highest rating are performing on this day.

Description of the process steps:

1. The data was collected from published venue of electronic music events in the Bay Area (https://19hz.info/eventlisting_BayArea.php), by using read_html into a table.
2. The data was cleaned by parsing out columns with performing artists, hosting clubs and the events dates and times. Results are exported into "output/event_artist_current.csv".
3. Using the obtained artists list, data about the number of followers and posted tracks was retrieved from SoundCloud.com for each artist, if found. The number of followers and posted tracks was merged with the data from published venue of planned events into a final data frame.
4. The final data frame was sorted in a descending order according to their number of followers, and included information about the hosting event/club, date and time, then and exported into pgAdmin house_cloud database. (all collected data is also exported to "output/current_venue_sorted_2.csv").
5. An app was constructed that is reading the pgAdmin database and presenting the results on html pages with three available routes: all found venue; venue found for this month, and venue found for next month.

Results and visualizations:

Number of followers on SoundCloud is representing a popularity of an artist and was used as the main sorting parameter. The number of posted tracks can give hints about the artists creativity and productivity, although can be a higher number for younger or starting artists to gain initial popularity.
The figure below shows the top 20 artists found, according to their number of followers:

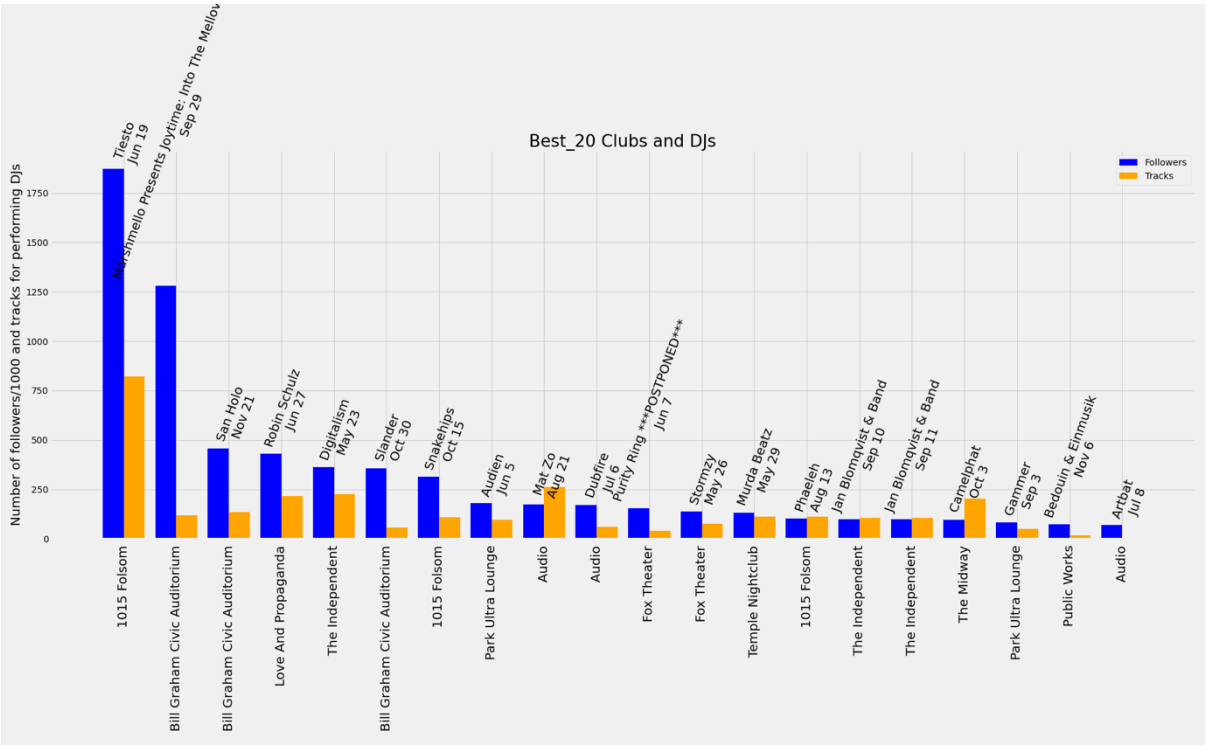


Fig. 1: Top 20 artists found from the current published venue.

As clearly seen, there is a large difference in the number of followers for the first two artists (Tiesto and Marshmello) due to their world-known level popularity. However, house music funs often are interested to avoid ‘too popular’ venues and stick more with an ‘underground’, slightly less popular style. The figure below shows a zoomed picture of top 20 artists without including the first two and extending into the others two:

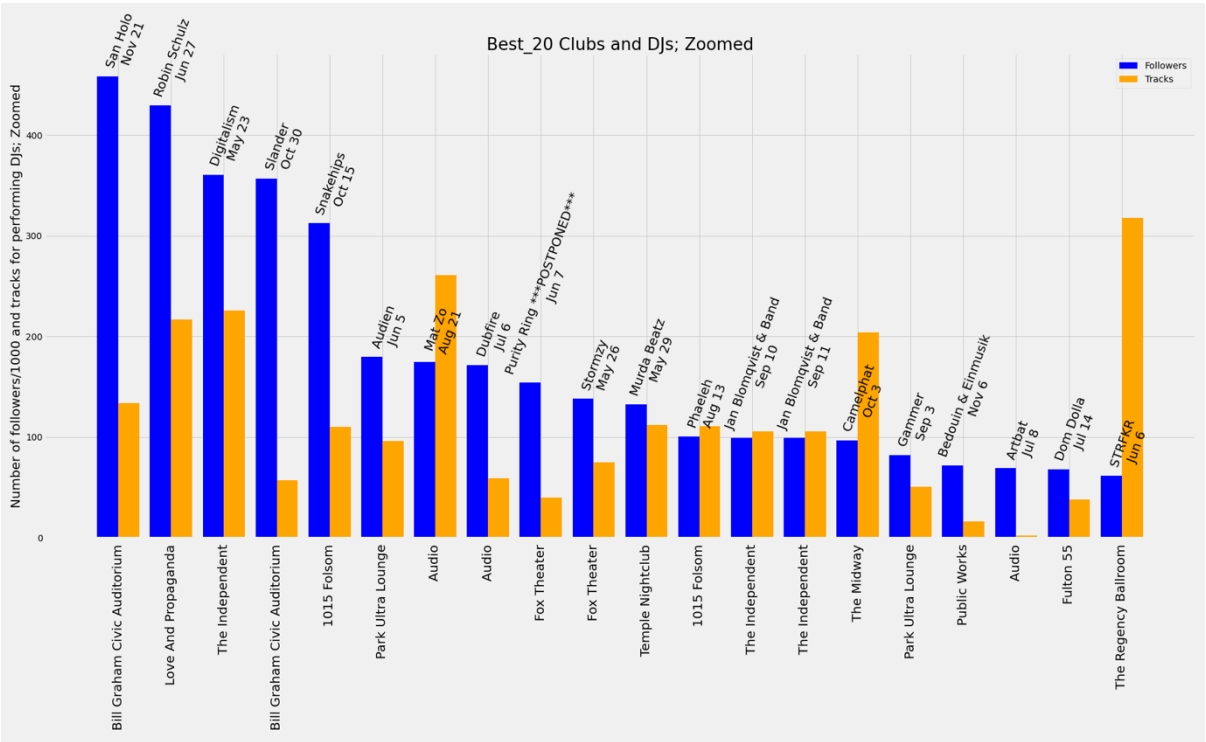
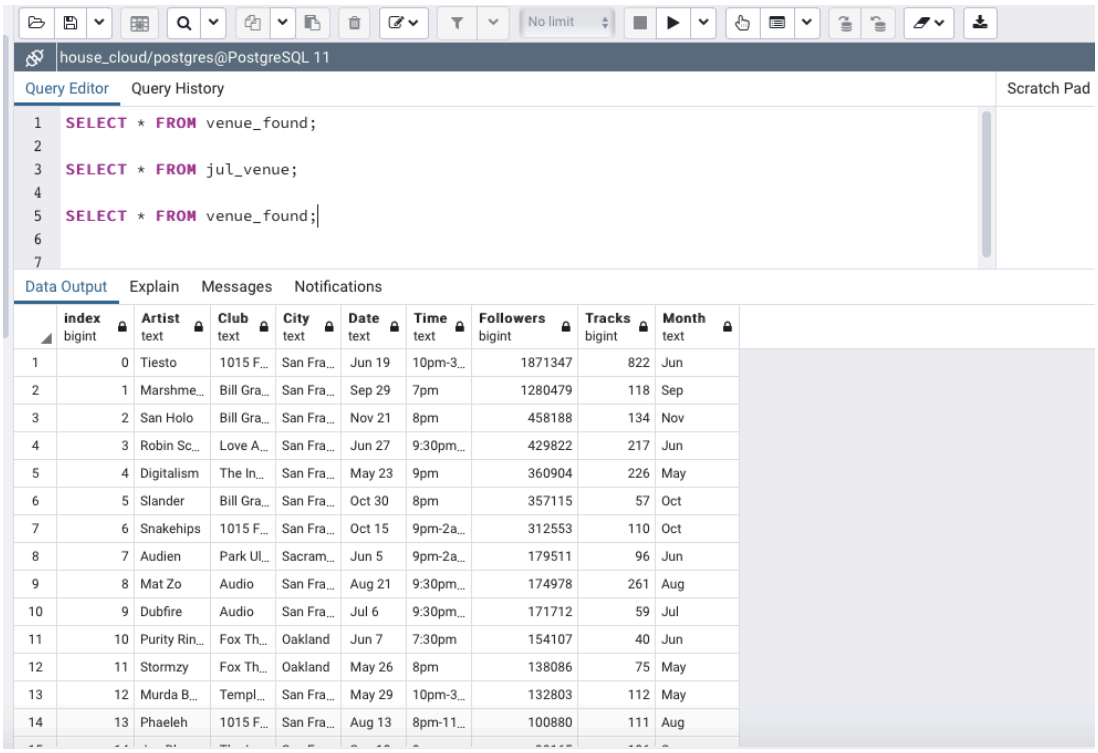


Fig. 2: Top 20 artists found from the current published venue, zoomed on an average popularity.
It is possible to update the zoom_f variable in order to zoom on another area (on Jupyter notebook).

Below shown an image of the database exported into pgAdmin:



	index bigint	Artist text	Club text	City text	Date text	Time text	Followers bigint	Tracks bigint	Month text
1	0	Tiesto	1015 F...	San Fra...	Jun 19	10pm-3...	1871347	822	Jun
2	1	Marshme...	Bill Gra...	San Fra...	Sep 29	7pm	1280479	118	Sep
3	2	San Holo	Bill Gra...	San Fra...	Nov 21	8pm	458188	134	Nov
4	3	Robin Sc...	Love A...	San Fra...	Jun 27	9:30pm...	429822	217	Jun
5	4	Digitalism	The In...	San Fra...	May 23	9pm	360904	226	May
6	5	Slander	Bill Gra...	San Fra...	Oct 30	8pm	357115	57	Oct
7	6	Snakehips	1015 F...	San Fra...	Oct 15	9pm-2a...	312553	110	Oct
8	7	Audien	Park Ul...	Sacram...	Jun 5	9pm-2a...	179511	96	Jun
9	8	Mat Zo	Audio	San Fra...	Aug 21	9:30pm...	174978	261	Aug
10	9	Dubfire	Audio	San Fra...	Jul 6	9:30pm...	171712	59	Jul
11	10	Purity Rin...	Fox Th...	Oakland	Jun 7	7:30pm	154107	40	Jun
12	11	Stormzy	Fox Th...	Oakland	May 26	8pm	138086	75	May
13	12	Murda B...	Templ...	San Fra...	May 29	10pm-3...	132803	112	May
14	13	Phaeleh	1015 F...	San Fra...	Aug 13	8pm-11...	100880	111	Aug

Fig. 3: All exported, found venue.

In order to present a venue for each month, 12 view tables were created in the pgAdmin as shown below and updating the month (including months where no venue was found; for future need when the search will be performed on a different time in a year):

```
CREATE VIEW may_venue AS
SELECT "Date", "Artist", "Club", "City", "Time", "Followers", "Tracks"
FROM venue_found
WHERE "Month" = 'May'
ORDER BY "Followers" DESC;
```

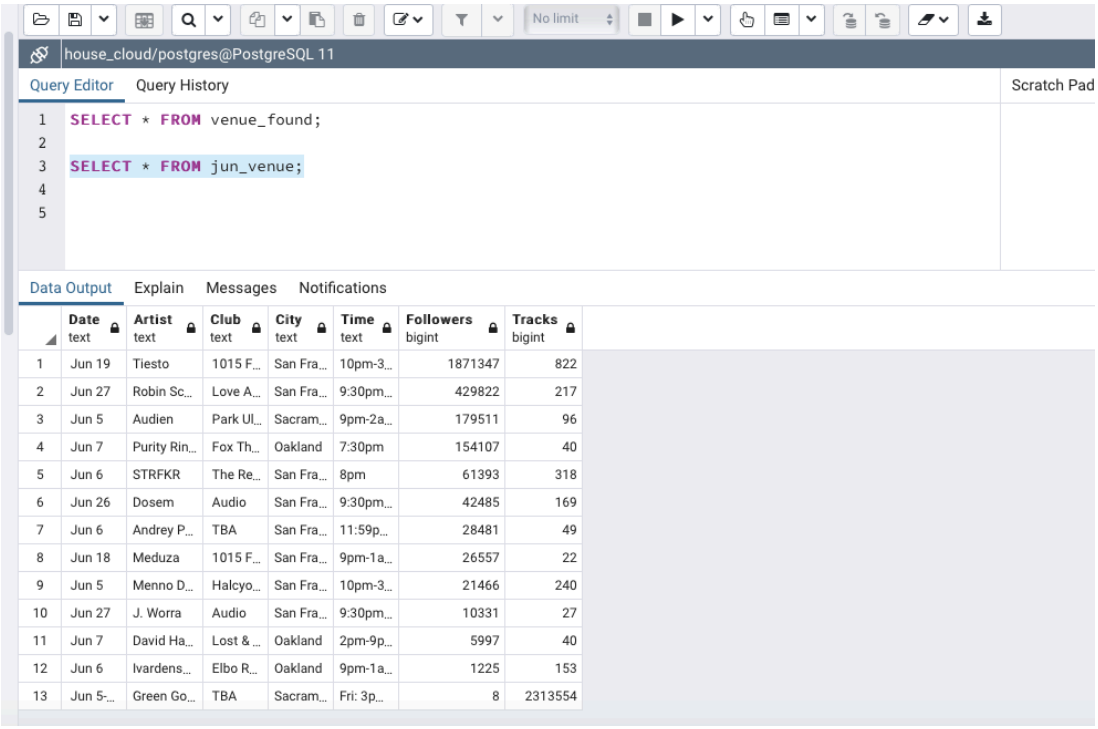


Fig. 4: Found June venue.

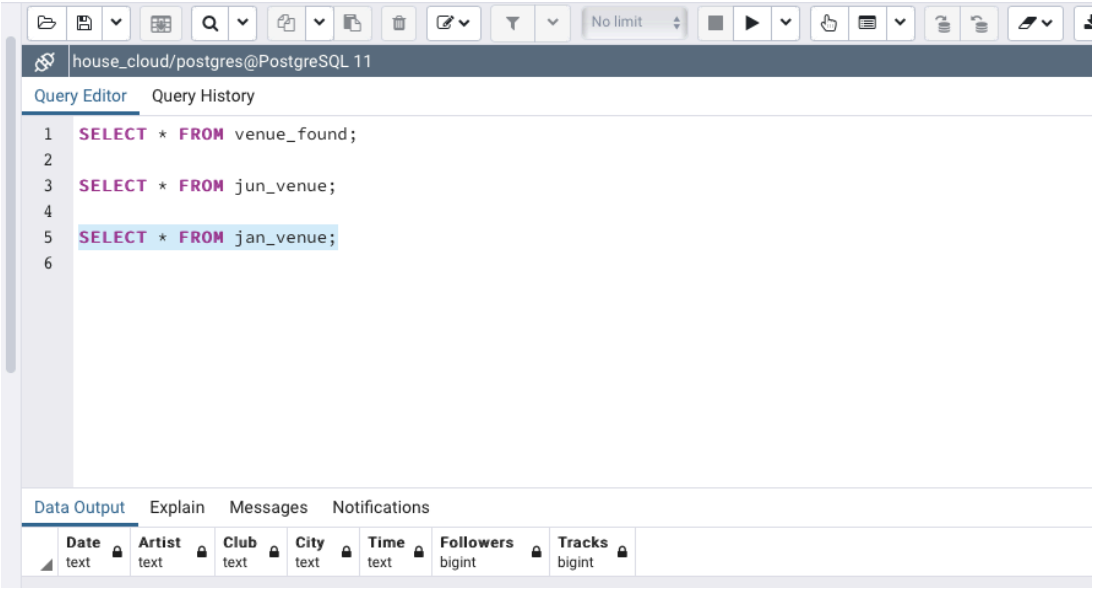


Fig. 4: Found January venue.

The final app.py reads the pgAdmin data base and constructed to show three available routes: all found venue (Fig. 6), venue found for this month (May, Fig. 7) and venue for the next month (June, Fig. 8). The database can be updated with a new data anytime, then just updating the "this" and "next" months names accordingly on the app.py:

```
venue_found = pd.read_sql_query('select * from venue_found', con=engine)
this_month_venue = pd.read_sql_query('select * from may_venue', con=engine)
next_month_venue = pd.read_sql_query('select * from jun_venue', con=engine)
```

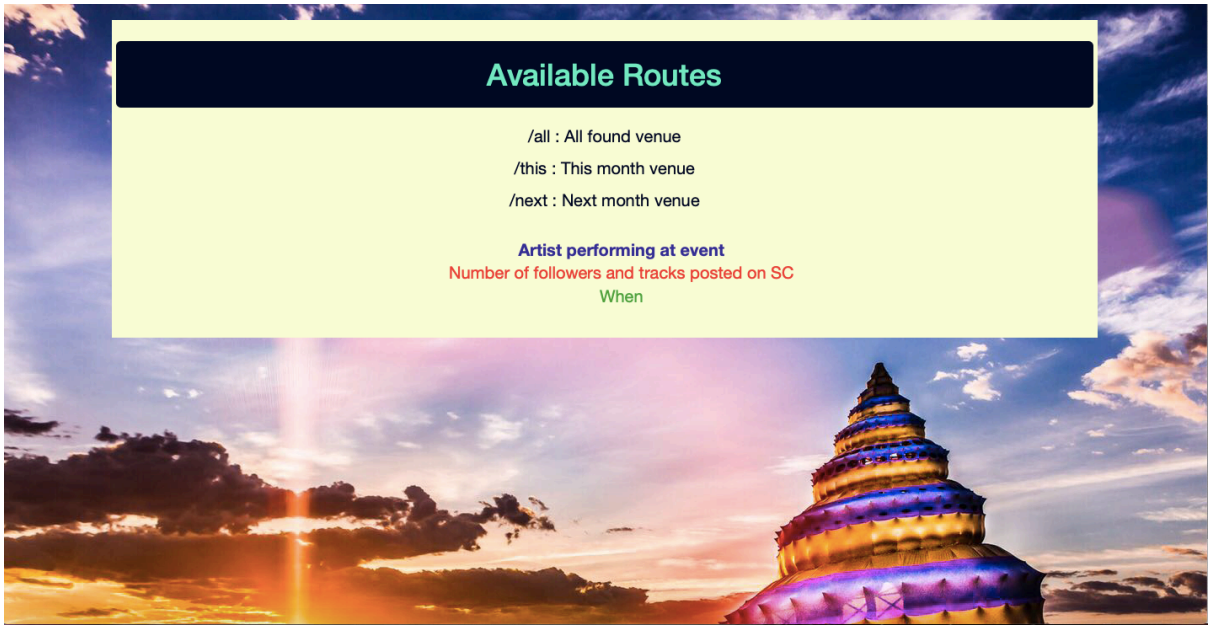


Fig. 5: Available routes on the app (/).

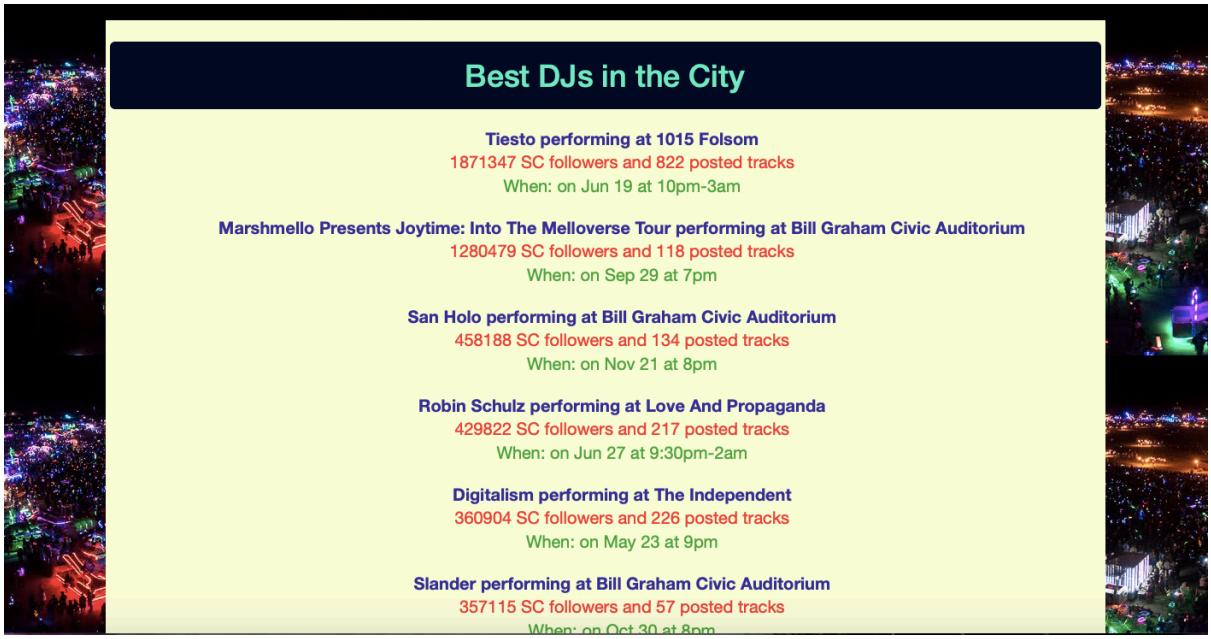


Fig. 6: All found venue route on the app (/all).

Fig. 7: Venue found for this month (/this).

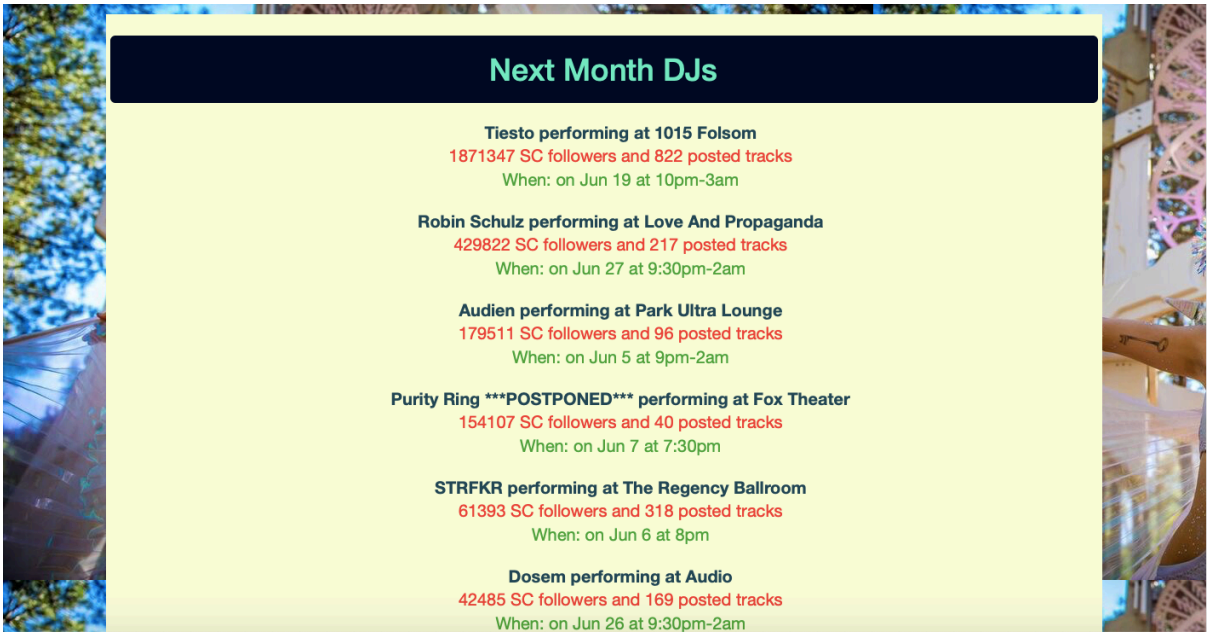


Fig. 8: Venue found for next month (/next).

Comments:

**** Due to COVID-19 published venue was significantly shorter than expected.
**** Username/Password for pgAdmin connection was removed for safety (please reach to me, if needed).

Data Sources:

1. https://19hz.info/eventlisting_BayArea.php
2. https://soundcloud.com