

**Big Data Project**

**Phase 3**

**Theme: Music**

Rajarshi Tarafdar

Sovan Ghosh

Sarah Choi

Adhirath Kapoor

Hoang Nguyen

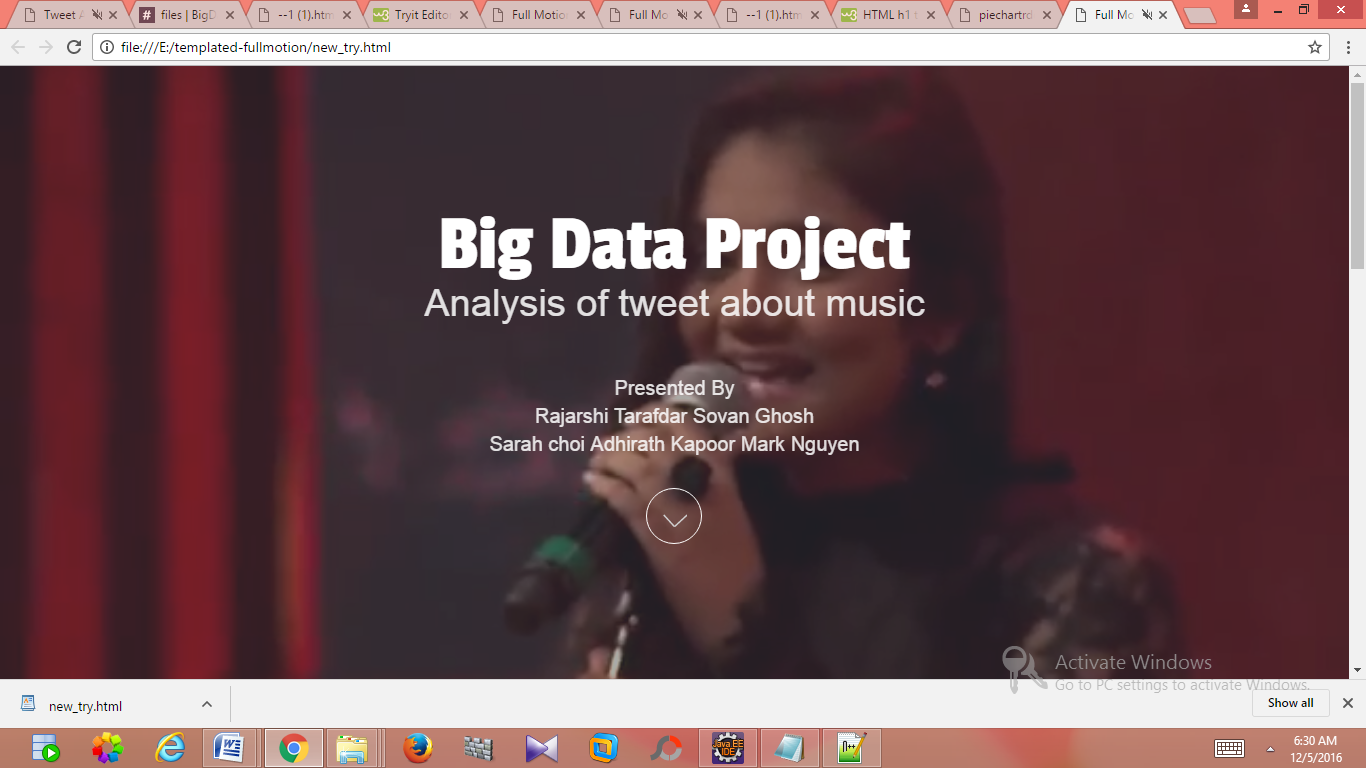
**Table of Content**

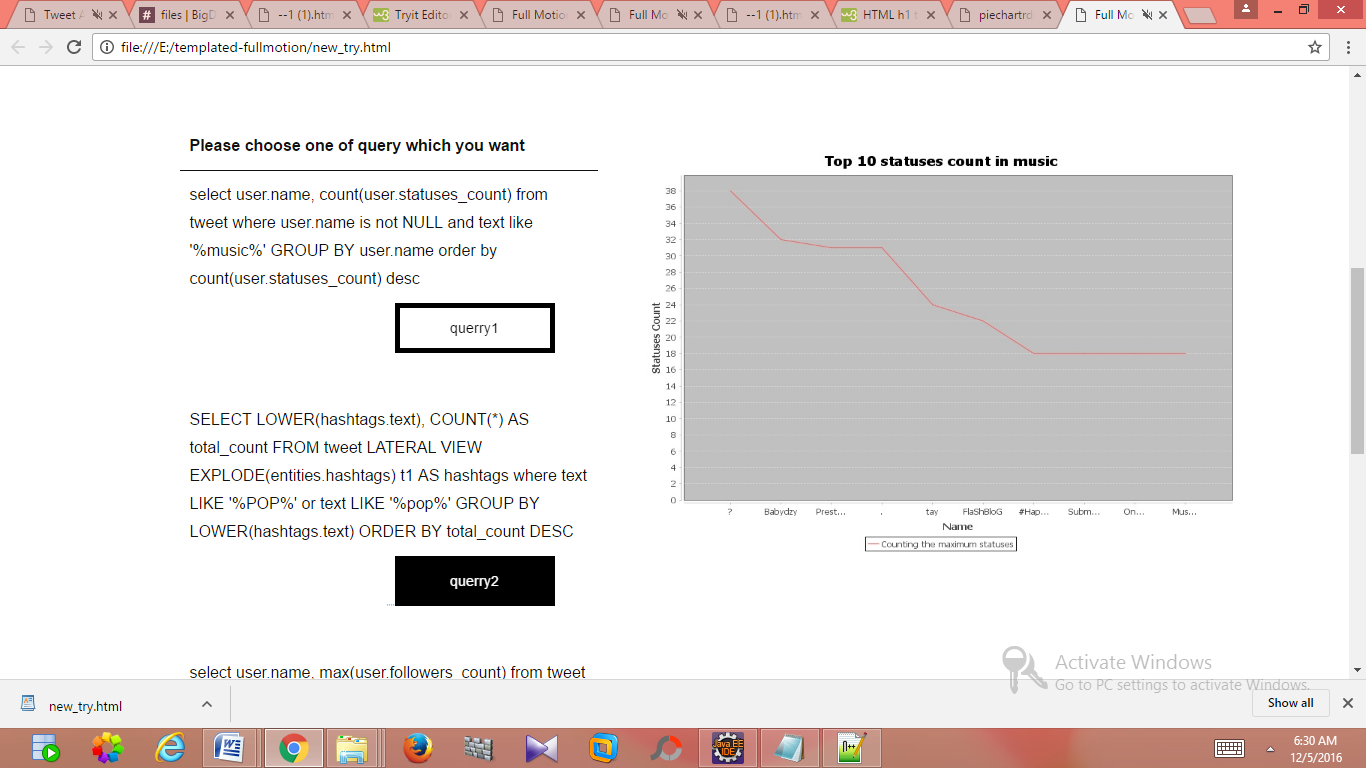
I. User Interface 3

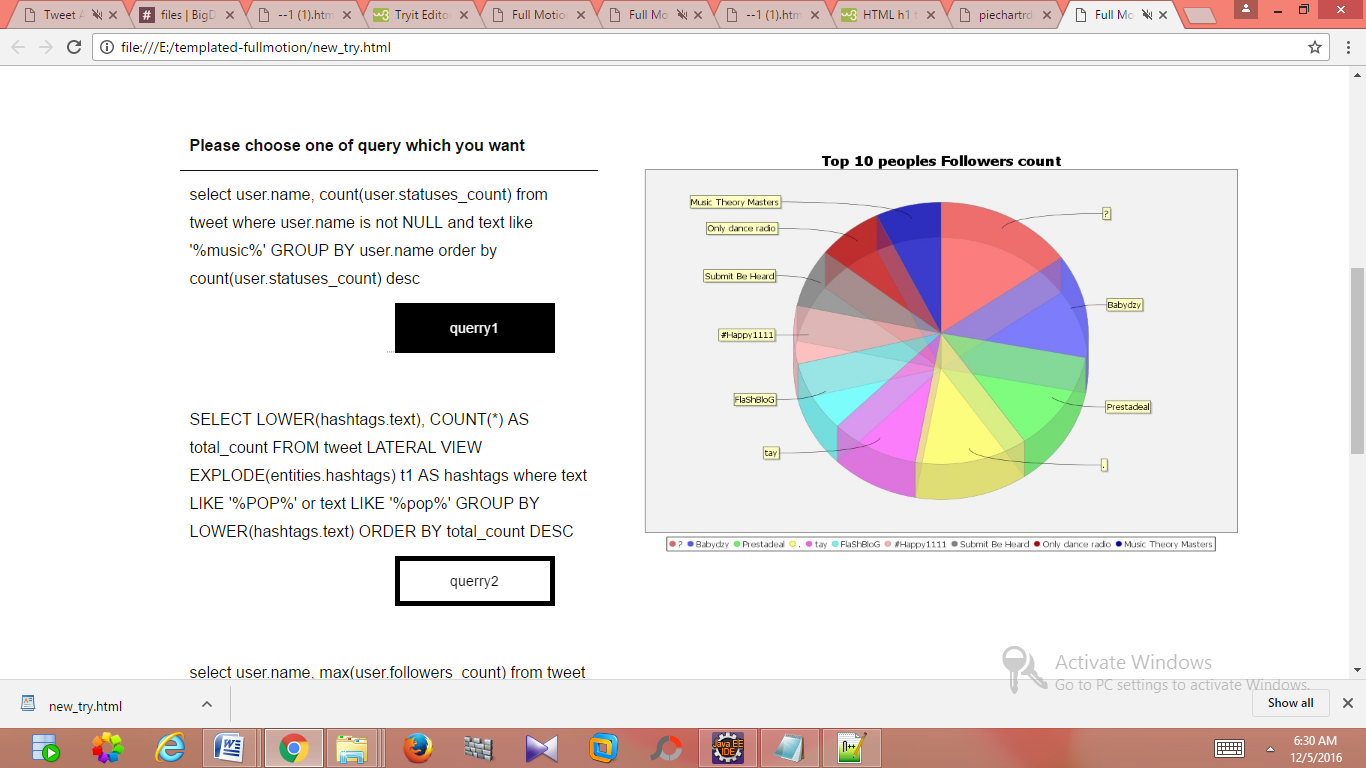
II. visualization

1. 1th query 5
2. 2th query 6
3. 3th query 7
4. 4th query 8
5. 5th query 9

# **User Interface**



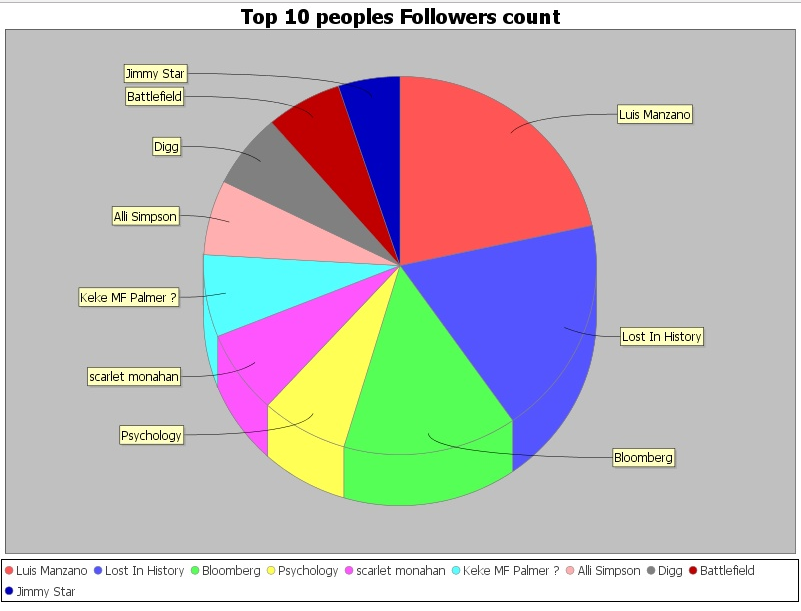




# **Visualization**

**DataFrames**

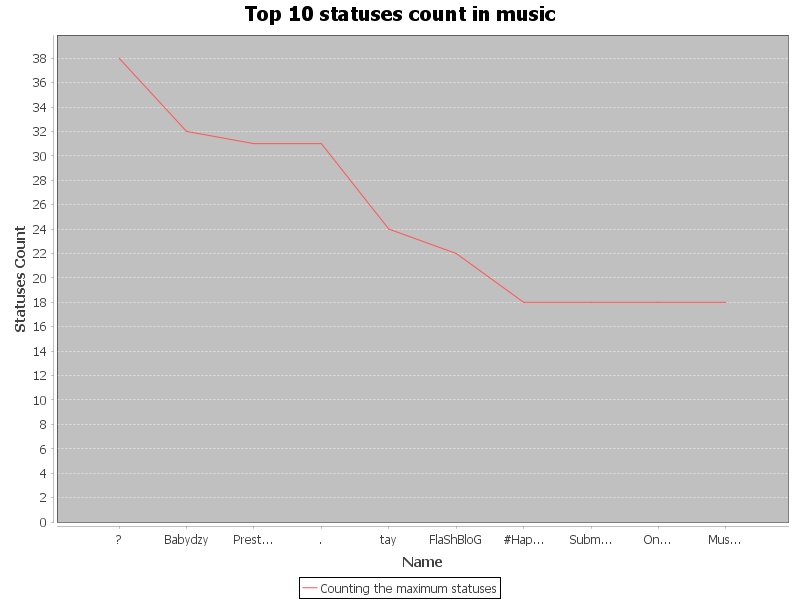
1. select user.name, max(user.followers\_count) from tweet where user.id IS NOT NULL group by user.name order by max(user.followers\_count) desc



**Explain**: get users that have the most follower

**Meaning:** get the most popular users with the most followers

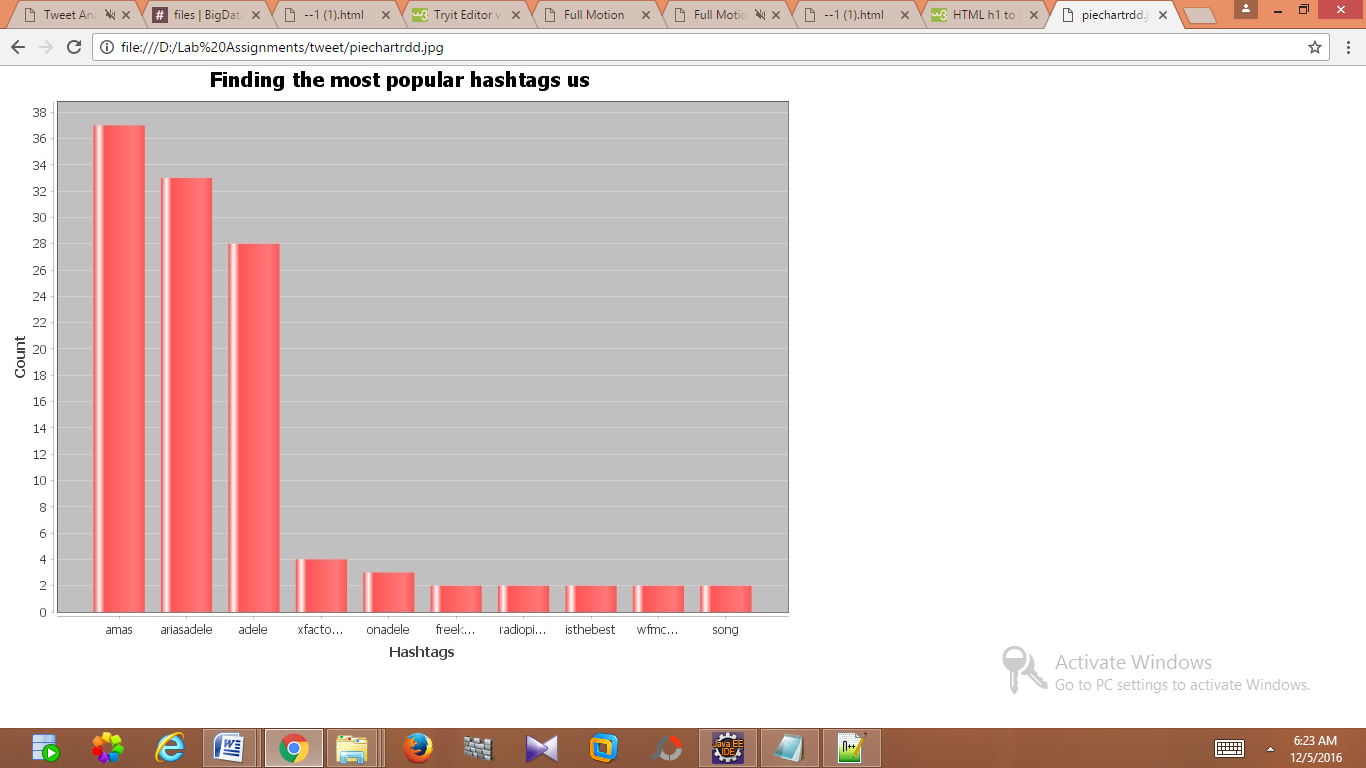
1. select user.name, count(user.statuses\_count) from tweet where user.name is not NULL and text like '%music%' GROUP BY user.name order by count(user.statuses\_count) desc



**Explain**: find the most popular user who tweeted the most

**Meaning**: we want to get top users that talk the most about music

1. SELECT LOWER(hashtags.text), COUNT(\*) AS total\_count FROM tweet LATERAL VIEW EXPLODE(entities.hashtags) t1 AS hashtags where text LIKE '%ADELE%' or text LIKE '%Adele%' GROUP BY LOWER(hashtags.text) ORDER BY total\_count DESC

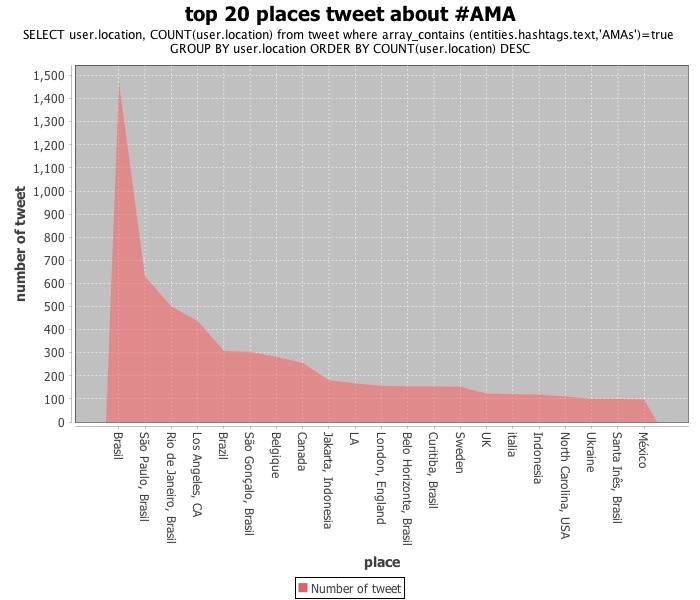


**Explain**: get the most popular hashtag used when tweet mentions Adele

**Meaning**: With most hashtag when the users talk about Adele so that we can use those hashtags in advertisement

**RDD**

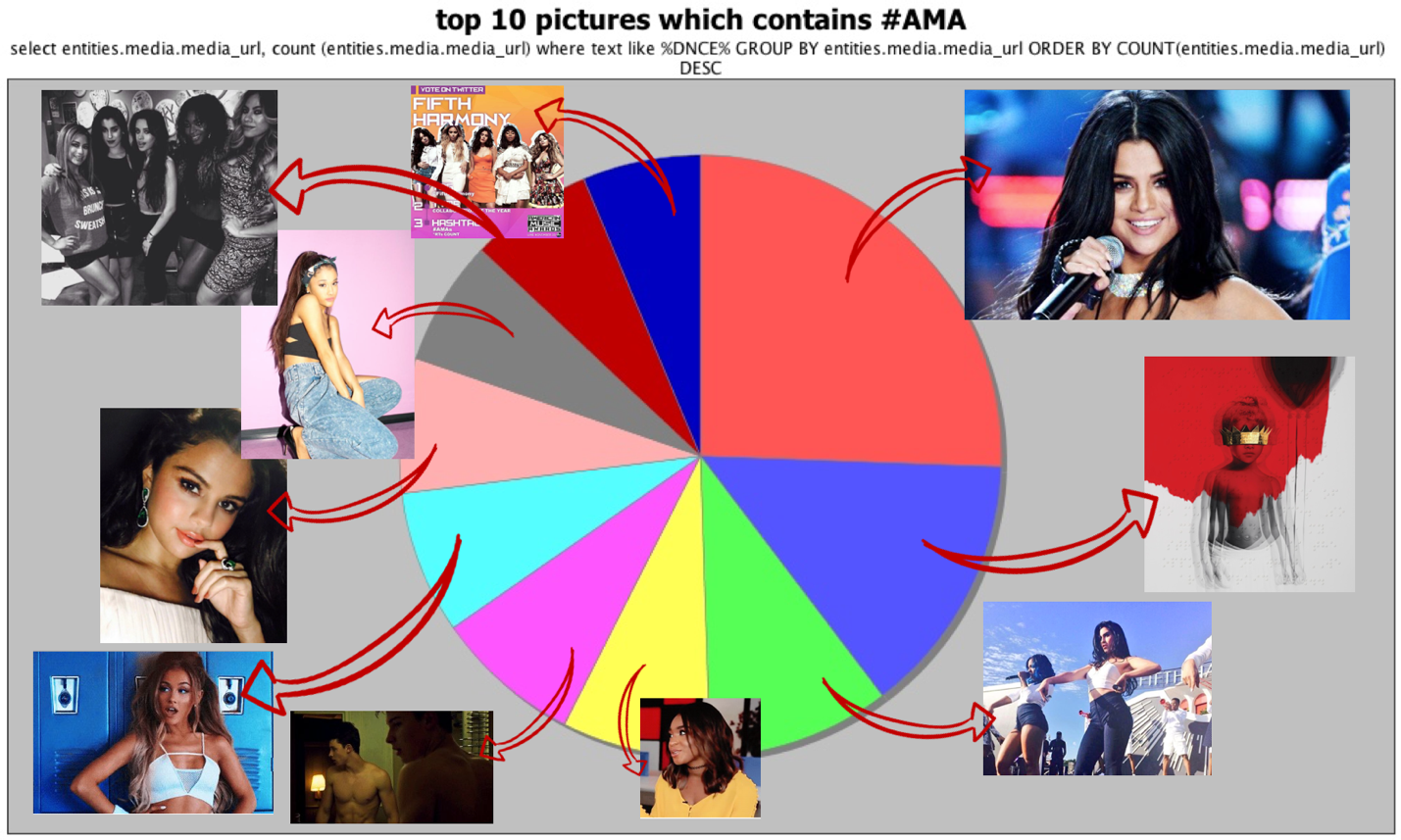
1. SELECT user.location, COUNT(user.location) from tweet where array\_contains (entities.hashtags.text,'AMAs')=true GROUP BY user.location ORDER BY COUNT(user.location) DESC



**Explain:** get top locations that have the most tweets with hashtag “AMAs” in them.

**Meaning:** With the top locations, we can know the popularity of AMA in different cities. We can see a lot of Brazil users are mentioning ‘AMAs’.

1. select entities.media.media\_url, count (entities.media.media\_url) where text like %DNCE% GROUP BY entities.media.media\_url ORDER BY COUNT(entities.media.media\_url) DESC



**Explain**: get the top popular photo urls that have the hashtag “AMA” in the tweet

**Meaning**: we know which pictures is used the most when the users use hashtag DNCE