

## ANALYSIS OF CELEBRITIES PROFILE

---

### Description:

This data set contains the data about celebrities' profiles of Twitter.

### Parameters:

**numeric\_id:** A unique id number for every profile.

**Verified:** Status of id verified or not.

**profile\_sidebar\_fill\_color:** The color for sidebar of profile.

**profile\_text\_color:** The color of Text for every profile.

**followers\_count:** Total number of followers of each celebrity.

**Protected:** Profiles are protected or not.

**Location:** Current location mentioned in the profile.

**profile\_background\_color:** Background color of the profile.

**utc\_offset:** The difference in hours and minutes from Universal Time (UTC) for their current location.

**statuses\_count:** Total status count of each celebrity.

**Description:** The status description of a person.

**friends\_count:** Total Friends of Celebrities

**profile\_link\_color:** The color of profile link.

**profile\_image\_url:** Image URL of the profile.

**Notifications:** The notification of the profile.

**profile\_background\_image\_url:** Background image url

**screen\_name**

**profile\_background\_tile**

**favourites\_count**

**name:** name of the celebrity

**url:** Profile URL.

**created\_at:** when the profile is created

**time\_zone:**

**profile\_sidebar\_border\_color:**

**following:** To whom they are following.

**gender**

### Problem and Solutions:

1. Find the lines that contains http. Or, Find the twitter profile image link for the accounts in which users uses an uploaded image as profile image not the default image.
2. Load the data from HDFS and find the total number of celebrities that have more than 80000 followers.
3. Perform the WordCount in the file and store the output into the HDFS.
4. Find the celebrity profile name whose profile is verified (i.e. true) and location is San Francisco.
5. Find the celebrity name who has the maximum followers.