**Queries**

**Query 1:**

Query1 = spark.sql(“SELECT substring( user.created\_at, 5, 3) as month, count(user.id) as count from tweets group by month”)

Query1.show()

**Query 2:**

Query2 = spark.sql(“SELECT count(\*) as count,user.name from tweets where user.name is not null group by user.name order bycount desc limit 10”)

Query2.show()

**Query 3:**

Query3 = spark.sql(“select count(\*) as count, q.text from ( select case when lower(text) like ‘%nike%’ then ‘nike’ when lower(text) like ‘%underarmour%’ then ‘underarmour’ when lower(text) like ‘%adidas%’ then ‘adidas’ when lower(text) like ‘%puma%’ then ‘puma’ when lower(text) like ‘%reebok%’ then ‘reebok’ else ‘different sports’ end as text from tweets)q group by q.text”)

Query3.show()

**Query 4:**

Query4 = spark.sql(“SELECT place.country,count(\*) AS count FROM tweets WHERE place.country IS NOT NULL GROUP BY place.country ORDER BY count DESC LIMIT 10”)

Query4.show()

**Query 5:**

Query5 = spark.sql(“SELECT substring(user.created\_at, 1, 3) as day,count(\*) as count from tweets group by day”)

Query5.show()

**Query 6:**

Query6 = spark.sql(“SELECT user.verified,user.screen\_name,user.followers\_count FROM tweets WHERE user.verified = false ORDER BY user.followers\_count DESC LIMIT 20”);

Query6.show()

**Query 7:**

Query7 = spark.sql(“SELECT user.screen\_name,text,retweeted\_status.retweet\_count FROM tweets ORDER BY retweeted\_status.retweet\_count DESC LIMIT 20”);

Query7.show()

**Query 8:**

Query8 = spark.sql(“SELECT user.location,count(text) as count FROM tweets WHERE place.country = ‘United States’ AND user.location is not null GROUP BY user.location ORDER BY count DESC LIMIT 10”);

Query8.show()

**Query 9:**

Q9 = spark.sql(“SELECT distinct id, “+

“CASE when user.verified LIKE ‘%true%’ THEN ‘VERIFIED ACCOUNT’”+

“when user.verified LIKE ‘%false%’ THEN ‘NON-VERIFIED ACCOUNT’”+

“END AS Verified from tweets where text is not null”);

Q9.createOrReplaceTempView(“acctVerify”);

var acctVerifydata = spark.sql(“SELECT Verified, Count(Verified) as Count from acctVerify where id is NOT NULL and Verified is not null group by Verified order by Count DESC”);

acctVerifydata.show()

**Query 10**

Query10 = spark.sql(“SELECT user.name, user.followers\_count, user.lang, text FROM tweets WHERE text like ‘%brand%’ order by user.followers\_count desc limit 15”)

Query10.show()