# Principles of Big Data Management

# CSEE-5540

# FALL-2015

Project Members:

1) Yatheeswara Reddy Pulicherla,
2) Shyam Raj Kinnera.

**Back End:**

MongoDB,

Java.

**Front End:**

Html,

Java script,

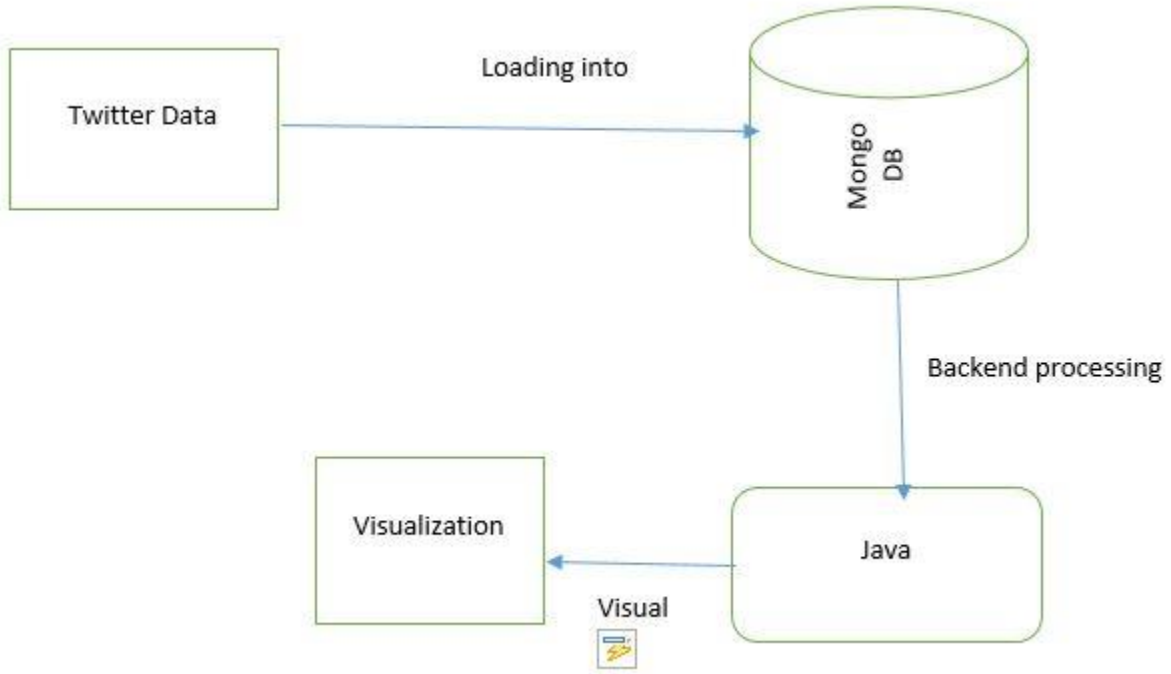
Css,

D3 Charts.

## Queries:

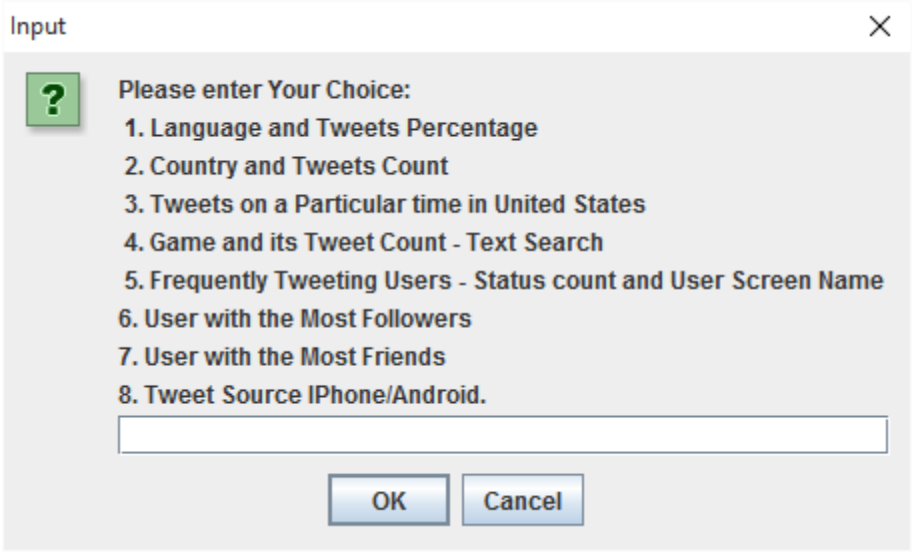
1) Language and Tweets Percentage.
2) Country and Tweets Count.
3) Tweets on a Particular time in United States.
4) Game and its Tweet Count - Text Search.
5) Frequently Tweeting Users - Status count and User Screen Name.
6) User with the Most Followers.
7) User with the Most Friends
8) Tweet Source IPhone/Android.

## Architecture Diagram:



## Implementation:

First we need to enter the choice for the query which we need to execute.



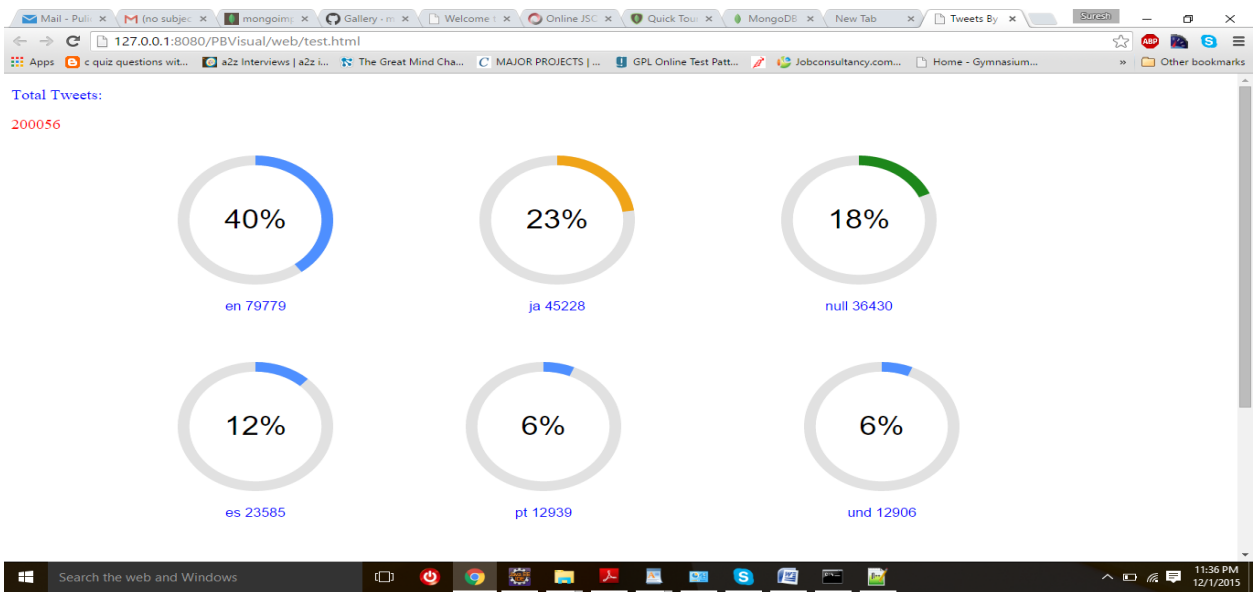

### 1) Language and Tweets Percentage:

The code for the query is:

```
try
            {           System.out.println("dgdjkdn");
                  FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res1.csv");
                  MongoClient mc = new MongoClient( "localhost" , 27017 );
                  DB db = mc.getDB( "tweets" );
                  DBCollection coll = db.getCollection("visual");
                  DBCursor cursor = coll.find();
                  BasicDBObject query = new BasicDBObject("i", 71);
                  long x=coll.getCount();
                  System.out.println("total records:"+x);
                  DBObject fields = new BasicDBObject("lang", 1);
                  fields.put("Language", 1);
                  //DBObject match = new BasicDBObject("$match", new
BasicDBObject());
                  DBObject project = new BasicDBObject("$project", fields );
                  DBObject groupFields = new BasicDBObject( "_id", "$lang");
                  groupFields.put("total", new BasicDBObject( "$sum", 1));
```

```
                    DBObject group = new BasicDBObject("$group", groupFields);
                    DBObject sort = new BasicDBObject("$sort", new
BasicDBObject("total", -1));
                    //List<DBObject> pipeline = Arrays.asList(match,project,
group, sort);
                    AggregationOutput output =coll.aggregate(project,
group,sort);
                    fw.append("Language");
                    fw.append(',');
                    fw.append("Count");
                    fw.append("\n");
                    for (DBObject result : output.results()) {
                        System.out.println(result);
                        fw.append((CharSequence) result.get("_id"));
                        fw.append(',');
                        fw.append(result.get("total").toString());
                        fw.append("\n");
                    }
                    fw.close();
                File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/test.html");
                Desktop.getDesktop().browse(f.toURI());
                }
                catch(Exception e)
                {
                        System.out.println("Exception:"+e);
                                        }
```

The result screen looks like as shown below.



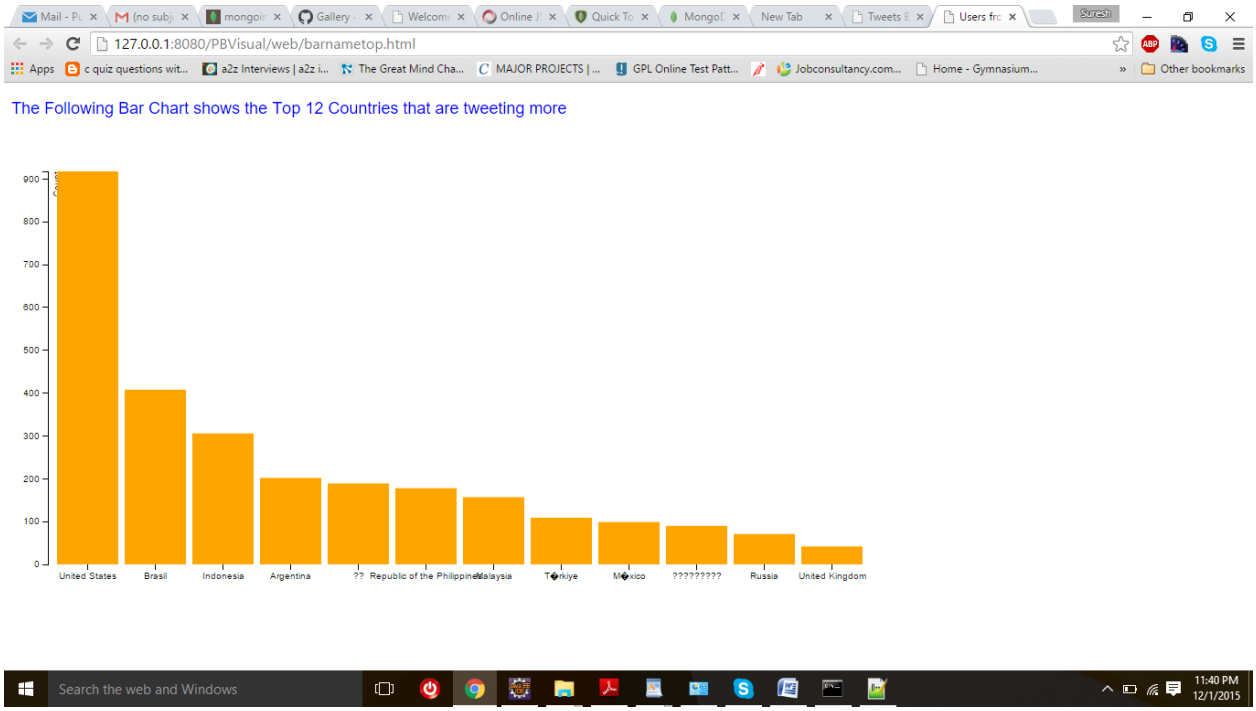

## 2) Country and Tweets Count:

# The code for the query is:

```
try
            {
                  FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res2.csv");
                  MongoClient mc = new MongoClient( "localhost" , 27017 );
                  DB db = mc.getDB( "tweets" );
                  DBCollection coll = db.getCollection("visual");
                  DBCursor cursor = coll.find();
                  long x=coll.getCount();
                  System.out.println("total records:"+x);
                  DBObject fields = new BasicDBObject("place.country", 1);
                  fields.put("_ id", 1);
                  //DBObject match = new BasicDBObject("$match", new
BasicDBObject());
                  DBObject project = new BasicDBObject("$project", fields );
                  DBObject groupFields = new BasicDBObject( "_id",
"$place.country");
                  groupFields.put("total", new BasicDBObject( "$sum", 1));
                  DBObject group = new BasicDBObject("$group", groupFields);
                  DBObject sort = new BasicDBObject("$sort", new
BasicDBObject("total", -1));
                  //List<DBObject> pipeline = Arrays.asList(match,project,
group, sort);
                  AggregationOutput output =coll.aggregate(project,
group,sort);
                  fw.append("Country");
                  fw.append(',');
                  fw.append("Count");
                  fw.append("\n");
                  int count=1;
                  for (DBObject result : output.results()) {
                        count++;
                  //    String t=result.get("_id").toString();
                        if(count>=15)
                              {
                              break;
                              }
                        if((result.get("_id")!=null))
                        {

                        System.out.println(result);
                    fw.append((CharSequence) result.get("_id"));
                    fw.append(',');
                    fw.append(result.get("total").toString());
                    fw.append("\n");
                      }

                  }
                  fw.close();
                  File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/barnametop.html");
                  Desktop.getDesktop().browse(f.toURI());
            }
            catch(Exception e)
```

```
		{
			System.out.println("Exception:"+e);
		}
```

The result screen is shown below,



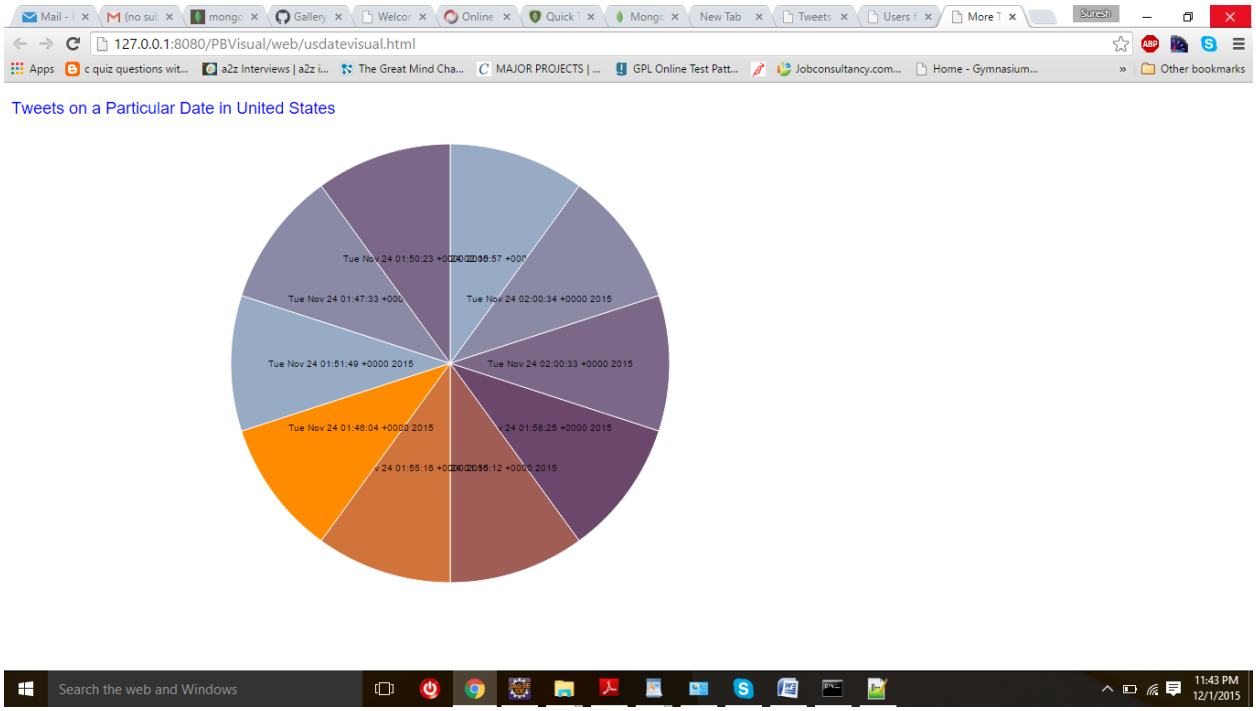

## **3)** Tweets on a Particular time in United States**:**

# The code for the query is:

```
try
		{
			FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res3.csv");
			MongoClient mc = new MongoClient( "localhost" , 27017 );
			DB db = mc.getDB( "tweets" );
			DBCollection coll = db.getCollection("visual");
			DBCursor cursor = coll.find();
			long x=coll.getCount();
			System.out.println("total records:"+x);
			DBObject fields = new BasicDBObject("created_at", 1);
			fields.put("_id", 1);
			DBObject match = new BasicDBObject("$match", new
BasicDBObject("place.country","United States"));
			DBObject project = new BasicDBObject("$project", fields );
```

```
                    DBObject groupFields = new BasicDBObject( "_id",
"$created_at");
                    groupFields.put("total", new BasicDBObject( "$sum", 1));
                    DBObject group = new BasicDBObject("$group", groupFields);
                    DBObject sort = new BasicDBObject("$sort", new
BasicDBObject("total", -1));
                    //List<DBObject> pipeline = Arrays.asList(match,project,
group, sort);
                    AggregationOutput output =coll.aggregate(match,project,
group,sort);
                    fw.append("date");
                    fw.append(',');
                    fw.append("Count");
                    fw.append("\n");
                    int count=1;
                    for (DBObject result : output.results()) {
                          count++;
                    //    String t=result.get("_id").toString();
                          if(count>11)
                                {
                                break;
                                }

                          System.out.println(result);
                        fw.append((CharSequence) result.get("_id"));
                        fw.append(',');
                        fw.append(result.get("total").toString());
                        fw.append("\n");
                    }
                    fw.close();
                    File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/usdatevisual.html");
                    Desktop.getDesktop().browse(f.toURI());
              }
              catch(Exception e)
              {
                    System.out.println("Exception:"+e);
              }
```

The result is shown as below,




# 4)Game and its Tweet Count - Text Search:

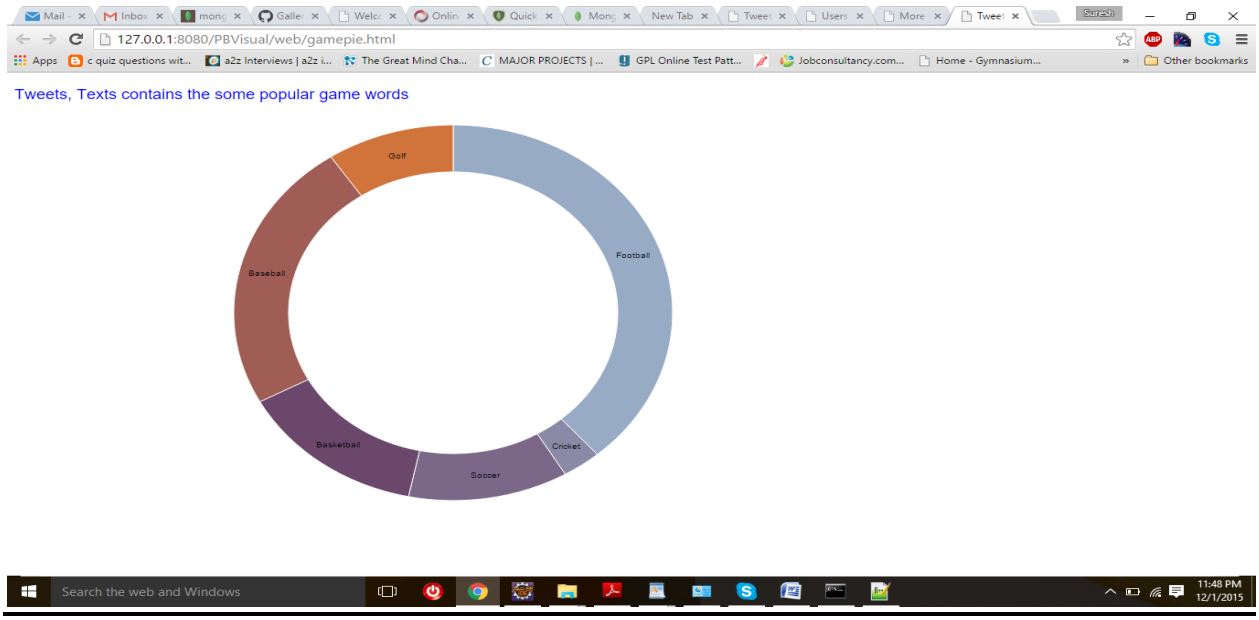
The code for the query is:

```
try
	{
		FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res4.csv");
		MongoClient mc = new MongoClient( "localhost" , 27017 );
		DB db = mc.getDB( "tweets" );
		DBCollection coll = db.getCollection("visual");
		long x=coll.getCount();
		System.out.println("total records:"+x);
		DBObject fields = new BasicDBObject("lang", 1);
		fields.put("_id", 1);
		Pattern fb = Pattern.compile(".*football.*",
Pattern.CASE_INSENSITIVE);
		BasicDBObject query = new BasicDBObject("text", fb);
		DBCursor cursor = coll.find(query);
		fw.append("Game");
		fw.append(',');
		fw.append("Count");
		fw.append("\n");
		int count = 0;
		while(cursor.hasNext()){
			count++;
			cursor.next();
		}
```

```
            System.out.println("Football "+count);
            fw.append("Football");
            fw.append(',');
            fw.append(count+"");
            fw.append("\n");

            Pattern cr = Pattern.compile(".*cricket.*",
Pattern.CASE_INSENSITIVE);
            BasicDBObject query1 = new BasicDBObject("text", cr);
            DBCursor cursor1 = coll.find(query1);
            count = 0;
            while(cursor1.hasNext()){
                  count++;
                  cursor1.next();
            }
            System.out.println("Cricket "+count);
            fw.append("Cricket");
            fw.append(',');
            fw.append(count+"");
            fw.append("\n");

            Pattern sc = Pattern.compile(".*soccer.*",
Pattern.CASE_INSENSITIVE);
            BasicDBObject query2 = new BasicDBObject("text", sc);
            DBCursor cursor2 = coll.find(query2);
            count = 0;
            while(cursor2.hasNext()){
                  count++;
                  cursor2.next();
            }
            System.out.println("Soccer "+count);
            fw.append("Soccer");
            fw.append(',');
            fw.append(count+"");
            fw.append("\n");

            Pattern bb = Pattern.compile(".*basketball.*",
Pattern.CASE_INSENSITIVE);
            BasicDBObject query3 = new BasicDBObject("text", bb);
            DBCursor cursor3 = coll.find(query3);
            count = 0;
            while(cursor3.hasNext()){
                  count++;
                  cursor3.next();
            }
            System.out.println("Basketball "+count);
            fw.append("Basketball");
            fw.append(',');
            fw.append(count+"");
            fw.append("\n");

            Pattern bbb = Pattern.compile(".*baseball.*",
Pattern.CASE_INSENSITIVE);
            BasicDBObject query4 = new BasicDBObject("text", bbb);
            DBCursor cursor4 = coll.find(query4);
            count = 0;
            while(cursor4.hasNext()){
```

```
                count++;
                cursor4.next();
            }
            System.out.println("Baseball "+count);
            fw.append("Baseball");
            fw.append(',');
            fw.append(count+"");
            fw.append("\n");

            Pattern gf = Pattern.compile(".*golf.*",
Pattern.CASE_INSENSITIVE);
            BasicDBObject query5 = new BasicDBObject("text", gf);
            DBCursor cursor5 = coll.find(query5);
            count = 0;
            while(cursor5.hasNext()){
                count++;
                cursor5.next();
            }
            System.out.println("Golf "+count);
            fw.append("Golf");
            fw.append(',');
            fw.append(count+"");
            fw.append("\n");
            fw.close();
            File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/gamepie.html");
            Desktop.getDesktop().browse(f.toURI());
        }
        catch(Exception e)
        {
            System.out.println("Exception:"+e);
        }
```

The result is shown as below,



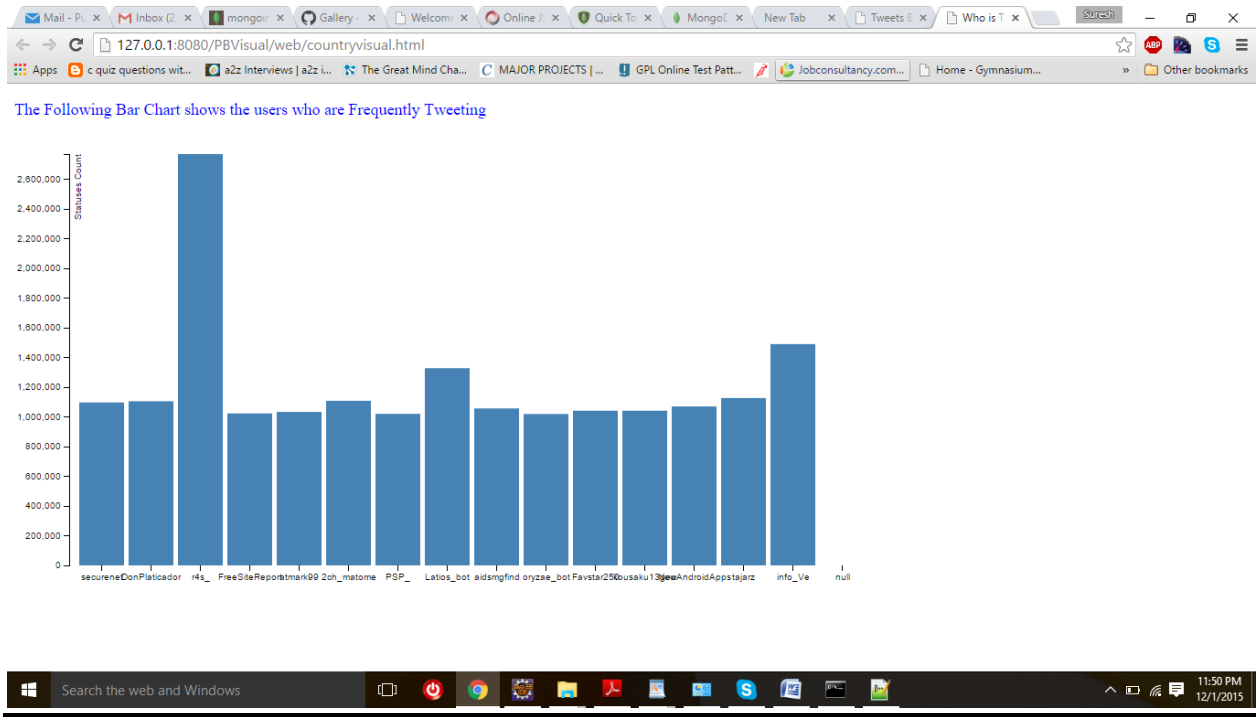
# 5)Frequently Tweeting Users - Status count and User Screen Name:

The code for the query is:

```
try
	{
		FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res5_1.csv");
		MongoClient mc = new MongoClient( "localhost" , 27017 );
		DB db = mc.getDB( "tweets" );
		DBCollection coll = db.getCollection("visual");
		DBCursor cursor = coll.find();
		long x=coll.getCount();
		System.out.println("total records:"+x);
		DBObject fields = new BasicDBObject("user.screen_name",
1).append("user.statuses_count", 1);
		fields.put("_id", 1);

		DBObject match = new BasicDBObject("$match", new
BasicDBObject("user.statuses_count",new BasicDBObject("$gt",1000000)));
		DBObject project = new BasicDBObject("$project", fields );

		Map<String, Object> dbObjIdMap = new HashMap<String, Object>();
		dbObjIdMap.put("name", "$user.screen_name");
		dbObjIdMap.put("count", "$user.statuses_count");
		DBObject groupFields = new BasicDBObject( "_id", new
BasicDBObject(dbObjIdMap));

		//DBObject groupFields = new BasicDBObject( "_id",
"$user.screen_name");
		//groupFields.put("count", new BasicDBObject( "count",
"$user.statuses_count"));
		DBObject group = new BasicDBObject("$group", groupFields);
		//DBObject sort = new BasicDBObject("$sort", new
BasicDBObject("total", -1));
		//List<DBObject> pipeline = Arrays.asList(match,project, group,
sort);
		AggregationOutput output =coll.aggregate(match,project, group);
		fw.append("Screen_name");
		fw.append(',');
		fw.append("Statuses");
		fw.append("\n");
		int count=1;
		for (DBObject result : output.results()) {

		  System.out.println(result);

		fw.append((CharSequence) result.get("$_id.name"));
		fw.append(',');
		fw.append((CharSequence) result.get("$_id.count"));
		fw.append("\n");
```

```
            }
            fw.close();
            File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/countryvisual.html");
            Desktop.getDesktop().browse(f.toURI());
      }
      catch(Exception e)
      {
            System.out.println("Exception:"+e);
      }
```

The result is shown as below,



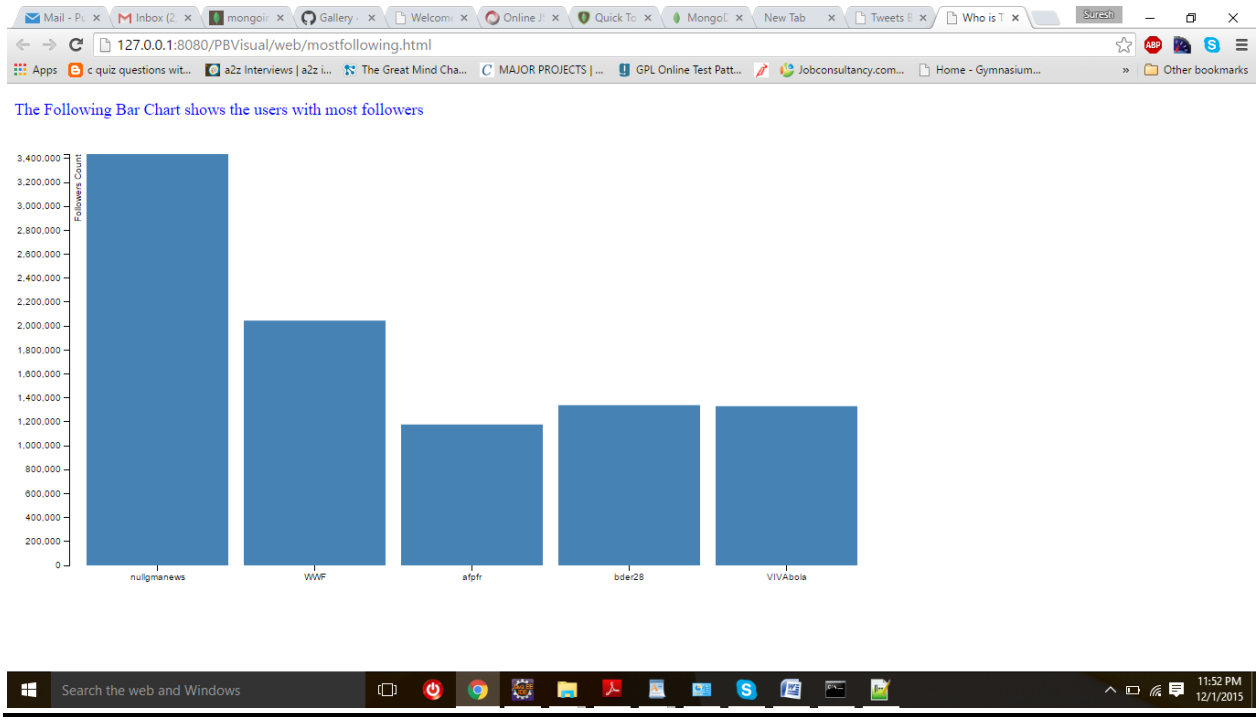

# 6)User with the Most Followers:

## The code for the query is:

```
try
            {
                  FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res6.csv");
                  MongoClient mc = new MongoClient( "localhost" , 27017 );
                  DB db = mc.getDB( "tweets" );
                  DBCollection coll = db.getCollection("visual");
                  DBCursor cursor = coll.find();
                  long x=coll.getCount();
                  System.out.println("total records:"+x);
```

```
                    DBObject fields = new BasicDBObject("user.screen_name",
1).append("user.followers_count", 1);
                    fields.put("_id", 1);

                    DBObject match = new BasicDBObject("$match", new
BasicDBObject("user.followers_count",new BasicDBObject("$gt",1000000)));
                    DBObject project = new BasicDBObject("$project", fields );

                    Map<String, Object> dbObjIdMap = new HashMap<String,
Object>();
                    dbObjIdMap.put("name", "$user.screen_name");
                    dbObjIdMap.put("count", "$user.followers_count");
                    DBObject groupFields = new BasicDBObject( "_id", new
BasicDBObject(dbObjIdMap));

                    //DBObject groupFields = new BasicDBObject( "_id",
"$user.screen_name");
                    //groupFields.put("count", new BasicDBObject( "count",
"$user.statuses_count"));
                    DBObject group = new BasicDBObject("$group", groupFields);
                    //DBObject sort = new BasicDBObject("$sort", new
BasicDBObject("total", -1));
                    //List<DBObject> pipeline = Arrays.asList(match,project,
group, sort);
                    AggregationOutput output =coll.aggregate(match,project,
group);
                    fw.append("Screen_name");
                    fw.append(',');
                    fw.append("followers");
                    fw.append("\n");
                    int count=1;
                    for (DBObject result : output.results()) {
                              if(count>5)
                                        break;
                      System.out.println(result);

                        fw.append((CharSequence) result.get("$_id.name"));
                        fw.append(',');
                        fw.append((CharSequence) result.get("$_id.count"));
                        fw.append("\n");
                          count++;

                    }
                    fw.close();
                    File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/mostfollowing.html");
                    //Desktop.getDesktop().browse(f.toURI());
               }
               catch(Exception e)
               {
                    System.out.println("Exception:"+e);
               }
```

The result is shown as below,



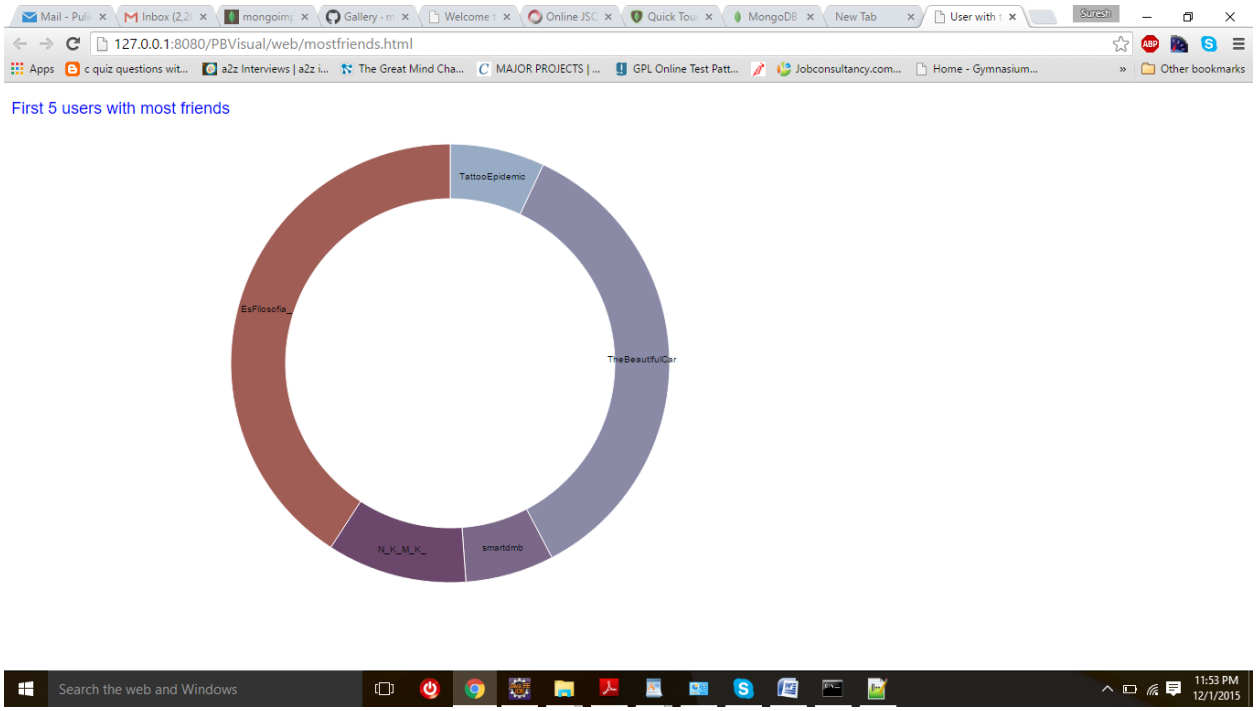

# 7)User with the Most Friends:

## The code for the query is:

```
try
            {
                  FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res7.csv");
                  MongoClient mc = new MongoClient( "localhost" , 27017 );
                  DB db = mc.getDB( "tweets" );
                  DBCollection coll = db.getCollection("visual");
                  DBCursor cursor = coll.find();
                  long x=coll.getCount();
                  System.out.println("total records:"+x);
                  DBObject fields = new BasicDBObject("user.screen_name",
1).append("user.friends_count", 1);
                  fields.put("_id", 1);

                  DBObject match = new BasicDBObject("$match", new
BasicDBObject("user.friends_count",new BasicDBObject("$gt",10000)));
                  DBObject project = new BasicDBObject("$project", fields );

                  Map<String, Object> dbObjIdMap = new HashMap<String,
Object>();
                  dbObjIdMap.put("name", "$user.screen_name");
                  dbObjIdMap.put("count", "$user.friends_count");
                  DBObject groupFields = new BasicDBObject( "_id", new
BasicDBObject(dbObjIdMap));
```

```
                    //DBObject groupFields = new BasicDBObject( "_id",
"$user.screen_name");
                    //groupFields.put("count", new BasicDBObject( "count",
"$user.statuses_count"));
                    DBObject group = new BasicDBObject("$group", groupFields);
                    //DBObject sort = new BasicDBObject("$sort", new
BasicDBObject("total", -1));
                    //List<DBObject> pipeline = Arrays.asList(match,project,
group, sort);
                    AggregationOutput output =coll.aggregate(match,project,
group);
                    fw.append("Screen_name");
                    fw.append(',');
                    fw.append("friends");
                    fw.append("\n");
                    int count=1;
                    for (DBObject result : output.results()) {
                              if(count>5)
                                   break;
                      System.out.println(result);

                        fw.append((CharSequence) result.get("$_id.name"));
                        fw.append(',');
                        fw.append((CharSequence) result.get("$_id.count"));
                        fw.append("\n");
                          count++;

                    }
                    fw.close();
                    File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/mostfriends.html");
                    //Desktop.getDesktop().browse(f.toURI());
               }
               catch(Exception e)
               {
                    System.out.println("Exception:"+e);
               }
```

The result is shown below,



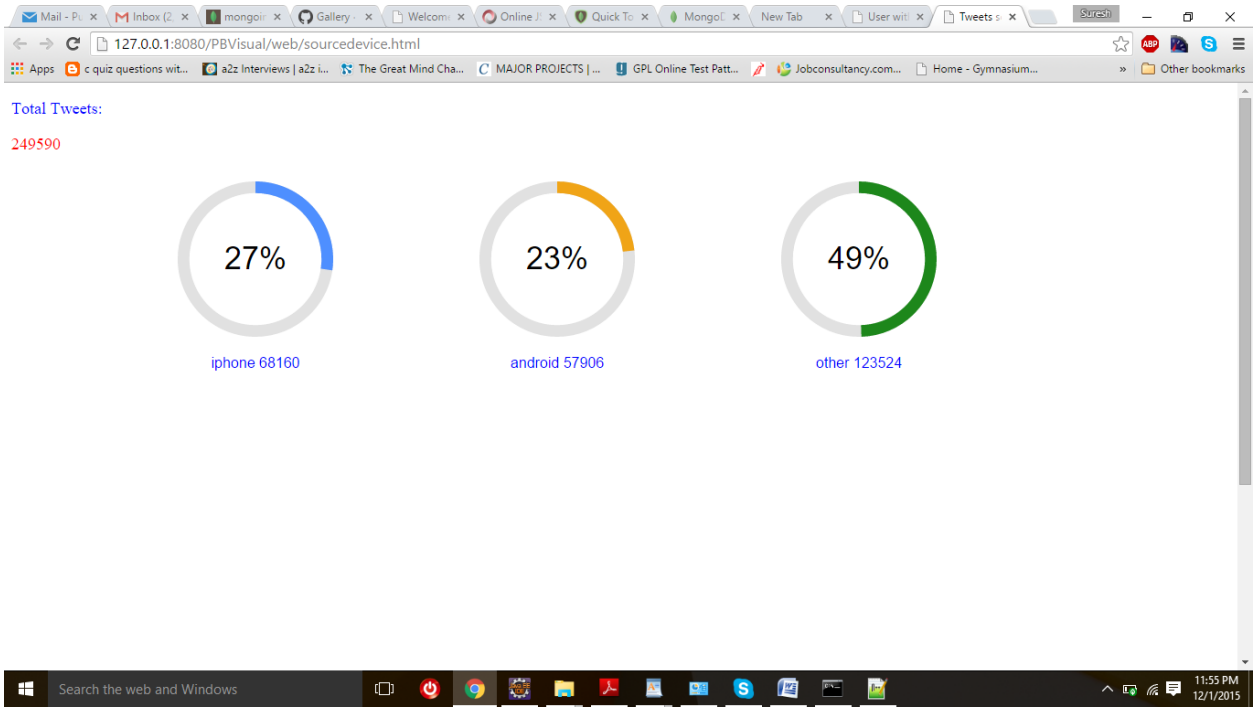

# 8)Tweet Source IPhone/Android:

The code for the query is:

```
try
            {
                  FileWriter fw= new
FileWriter("C:/Users/yarra/workspace/PBVisual/WebContent/web/res8.csv");
                  MongoClient mc = new MongoClient( "localhost" , 27017 );
                  DB db = mc.getDB( "tweets" );
                  DBCollection coll = db.getCollection("visual");
                  long x=coll.getCount();
                  System.out.println("total records:"+x);
                  DBObject fields = new BasicDBObject("lang", 1);
                  fields.put("_id", 1);
                  Pattern fb = Pattern.compile(".*iphone.*",
Pattern.CASE_INSENSITIVE);
                  BasicDBObject query = new BasicDBObject("source", fb);
                  DBCursor cursor = coll.find(query);
                  fw.append("device");
                  fw.append(',');
                  fw.append("Count");
                  fw.append("\n");
                  int count = 0;
                  while(cursor.hasNext()){
                        count++;
                        cursor.next();
                  }
```

```
                    System.out.println("iphone "+count);
                    fw.append("iphone");
                    fw.append(',');
                    fw.append(count+"");
                    fw.append("\n");

                    Pattern cr = Pattern.compile(".*android.*",
Pattern.CASE_INSENSITIVE);
                    BasicDBObject query1 = new BasicDBObject("source", cr);
                    DBCursor cursor1 = coll.find(query1);
                    count = 0;
                    while(cursor1.hasNext()){
                          count++;
                          cursor1.next();
                    }
                    System.out.println("android "+count);
                    fw.append("android");
                    fw.append(',');
                    fw.append(count+"");
                    fw.append("\n");

                    fw.append("other");
                    fw.append(',');
                    fw.append("123524");
                    fw.append("\n");

                    fw.close();
                    File f=new
File("C:/Users/yarra/workspace/PBVisual/WebContent/web/sourcedevice.html");
//                  Desktop.getDesktop().browse(f.toURI());
              }
              catch(Exception e)
              {
                    System.out.println("Exception:"+e);
              }
```

The result is shown as follows




## Deployment:

Bluemix url: http://pbdtwitterproject.mybluemix.net/web/index.html

Github link: https://github.com/YATHEESHWAR/PB