Azure Data Lake Assignment

Run these three scripts in Azure Data Lake Analytics

Super Bowl 50 Processing

Parse Super Bowl Tweets

```
REFERENCE ASSEMBLY [Newtonsoft.Json];
REFERENCE ASSEMBLY [TwitterProcessor];

@input =
        EXTRACT JsonString string
        FROM @"/superbowl50small/superbowl50small.txt"
        USING Extractors.Text(rowDelimiter: "\n", encoding: Encoding.UTF8, delimiter: '\b', quoting: fal
se);



@jsonExtracted =
        PROCESS @input
        PRODUCE id_string string,
                        tweet string,
                        created_at string,
                        favorited string,
                        retweeted string,
                        timestampMs string,
                        lang string,
                        user_id string,
                        user_location string,
                        friends_count string,
                        screen_name string,
                        name string,
                        time_zone string,
                        favorites_count string,
```

```
                                    retweet_count string
            USING new TwitterProcessor.TwitterJsonProcessor();

@processed =
            SELECT *
            FROM @jsonExtracted
            WHERE !String.IsNullOrWhiteSpace(created_at) AND !String.IsNullOrWhiteSpace(tweet);

OUTPUT @processed
TO "/output/<yourfisrtname>/superbowl50/superbowl50smallout.tsv"
USING Outputters.Tsv(Encoding.UTF8);
```

Super Bowl Tweet Summaries

```
@t = EXTRACT
         id string
        , text string
        , createdAt string
        , favorited string
        , retweeted string
        , timestampMs string
        , lang string
        , userId string
        , userLocation string
        , friendsCount string
        , screenName string
        , name string
        , timeZone string
        , favoritesCount string
        , retweetCount string
  FROM "/output/<yourfisrtname>/superbowl50/superbowl50smallout.tsv"
USING Extractors.Tsv(silent:true);

@res1 =
```

```
        SELECT lang,
                COUNT( * ) AS [tweet count]
        FROM @t
        GROUP BY lang;

OUTPUT @res1
TO "/output/<yourfisrtname>/superbowl50/superbowl50tweetsBYlangsmallOut.tsv"
ORDER BY [tweet count] DESC
USING Outputters.Tsv();

@res2 =
        SELECT userLocation,
                COUNT( * ) AS [tweet count]
        FROM @t
        GROUP BY userLocation;

OUTPUT @res2
TO "/output/<yourfisrtname>/superbowl50/superbowl50tweetsBYuserLocationsmallOut.tsv"
ORDER BY [tweet count] DESC
USING Outputters.Tsv();

@res3 =
        SELECT timeZone,
                COUNT( * ) AS [tweet count]
        FROM @t
        GROUP BY timeZone;

OUTPUT @res3
TO "/output/<yourfisrtname>/superbowl50/superbowl50tweetsBYtimeZonesmallOut.tsv"
ORDER BY [tweet count] DESC
USING Outputters.Tsv();
```

Super Bowl Tweet Detail

```
@t = EXTRACT
         id string
        , text string
        , createdAt string
        , favorited string
        , retweeted string
        , timestampMs string
        , lang string
        , userId string
        , userLocation string
        , friendsCount string
        , screenName string
        , name string
        , timeZone string
        , favoritesCount string
        , retweetCount string
  FROM "/output/<yourfisrtname>/superbowl50/superbowl50smallout.tsv"
USING Extractors.Tsv(silent:true);

@res =
              SELECT  id
                                    , screenName
                                    , text
                                    , lang
                                    , userLocation
                                    , timeZone
              FROM @t;

OUTPUT @res
TO "/output/<yourfisrtname>/superbowl50/superbowl50detailtweetssmallout.tsv"
USING Outputters.Tsv();
```

Home Work

After running these U-SQL scripts please create one more that finds the **screenNames** with the maximum **friendsCount**.

Submit your U-SQL script and the 3 screenNames and their friend counts for those with more than 9000 friends