Human Behaviour Analysis Using Intelligent Big Data Analytics

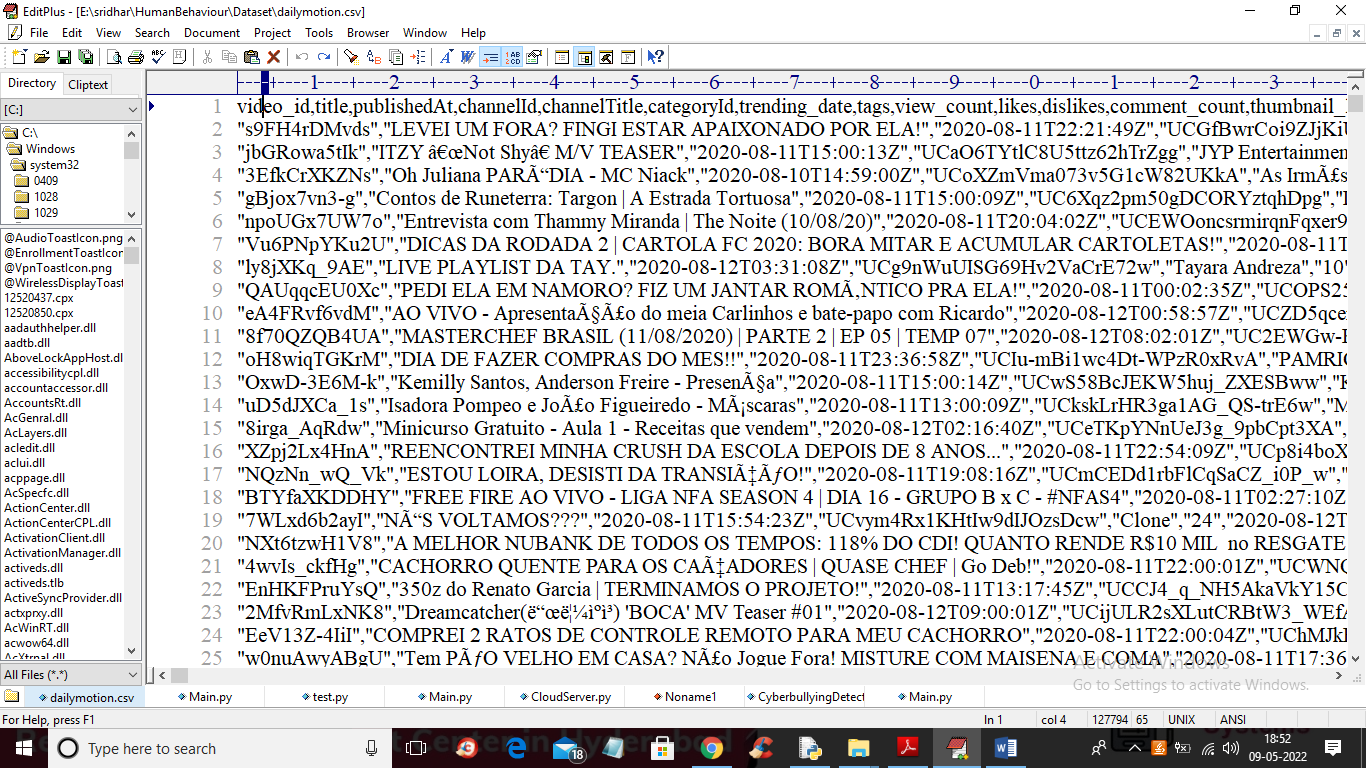
Now-a-days almost all peoples are using social media to express their views and by analysing this view we can predict person behaviour as their view often describe their personality but this social media contains reviews as TWEETS, POSTS in unstructured format and everyday this unstructured data gather in terabytes and if we want to extract meaningful information such as famous brand, powerful leader, most trending entertainment then this terabytes data processing may take huge time with traditional algorithms so author of this paper employing parallel processing techniques called SPARK.

SPARK will create multiple threads and then distribute work between those thread to perform task parallelly and then send result back to SPARK. All existing algorithms works on single thread but SPARK will distribute works in multiple threads so its paralleling processing will be faster and suitable for BIG DATA applications.

In propose paper author using HIVE, SPARK and HADOOP where HIVE and HADOOP will store data and SPARK will read and process that data. In propose work author gathering reviews from DAILYMOTION website as CSV file and then extracting useful information such as MOST TALK COUNTRIES with many LIKES and then extracting LIKES, VIEW and COMMENTS from so many categories called FASHION, ENTERTAINMENT, NEWS etc.

Author has compare execution time of SPARK processing and without spark processing and this experiment proves that SPARK is faster than traditional single thread processing.

We are using below CSV dataset of DAILYMOTION website to extract usefull information



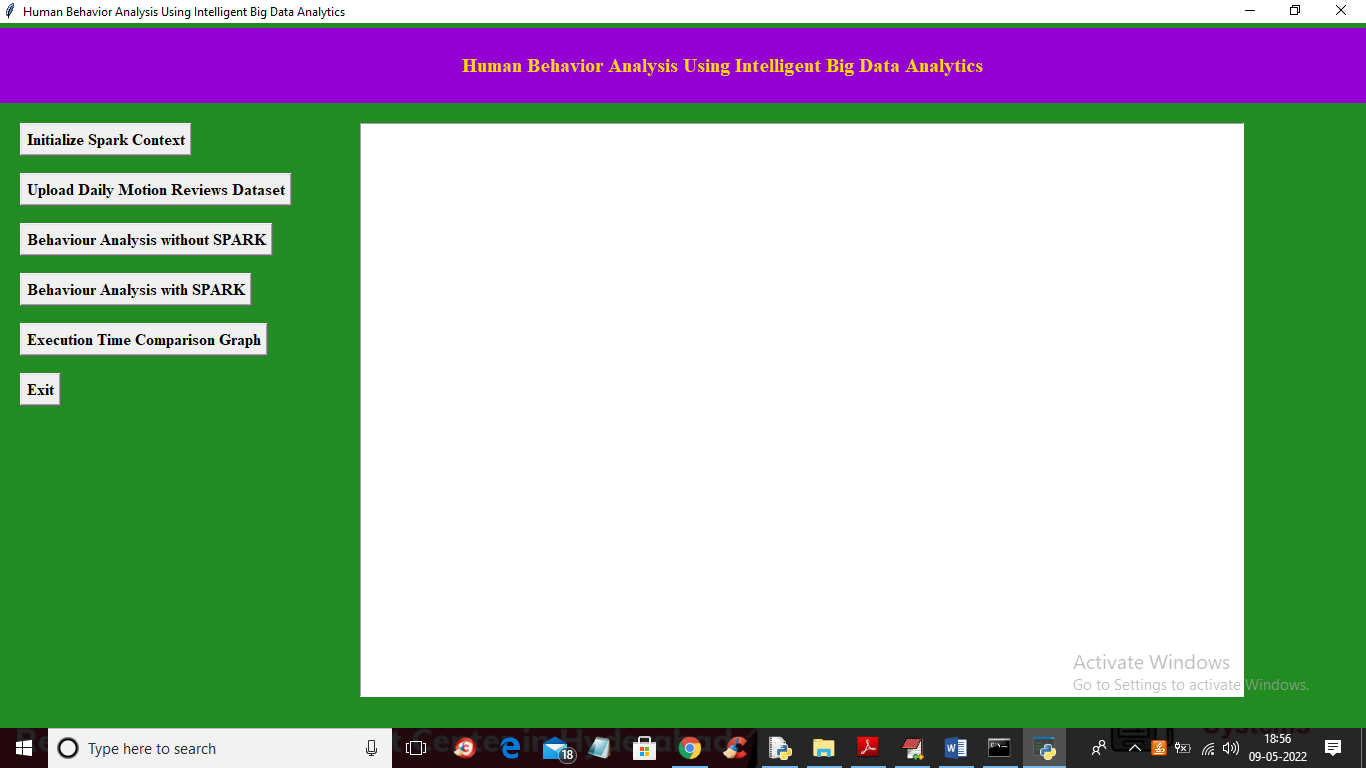
In above dataset screen first row contains dataset column names and remaining are the dataset values.

Modules Information

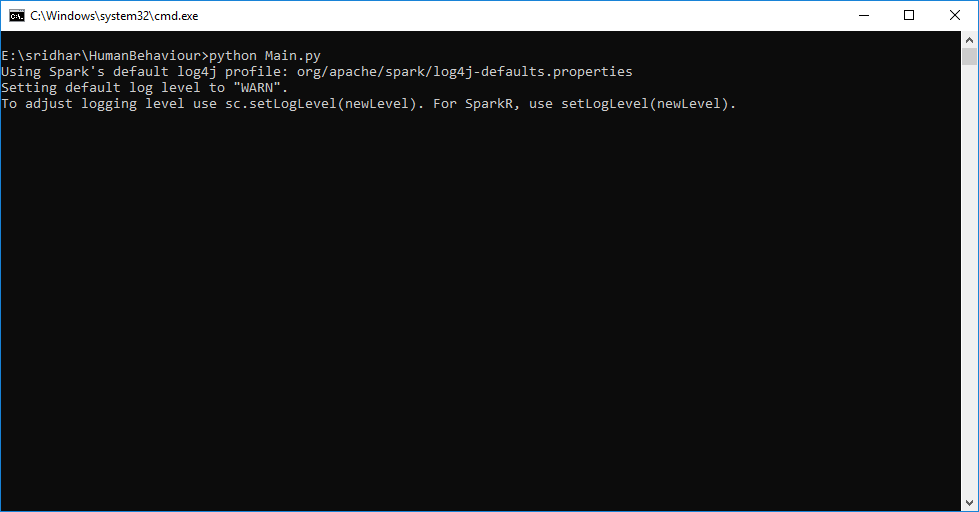
1. Initialize Spark Context: using this module we will initialize SPARK CONTEXT for parallel processing
2. Upload Daily Motion Reviews Dataset: using this module we will upload dataset file path to application
3. Behaviour Analysis without SPARK: using this module we will analyse human behaviour such as their LIKES, DISLIKES from their reviews without using SPARK technology and then capture its execution time
4. Behaviour Analysis with SPARK: using this module we will perform same task of behaviour analysis by using SPARK technology and then capture its execution time
5. Execution Time Comparison Graph: using this module we will plot execution time comparison between without and with SPARK processing

SCREEN SHOTS

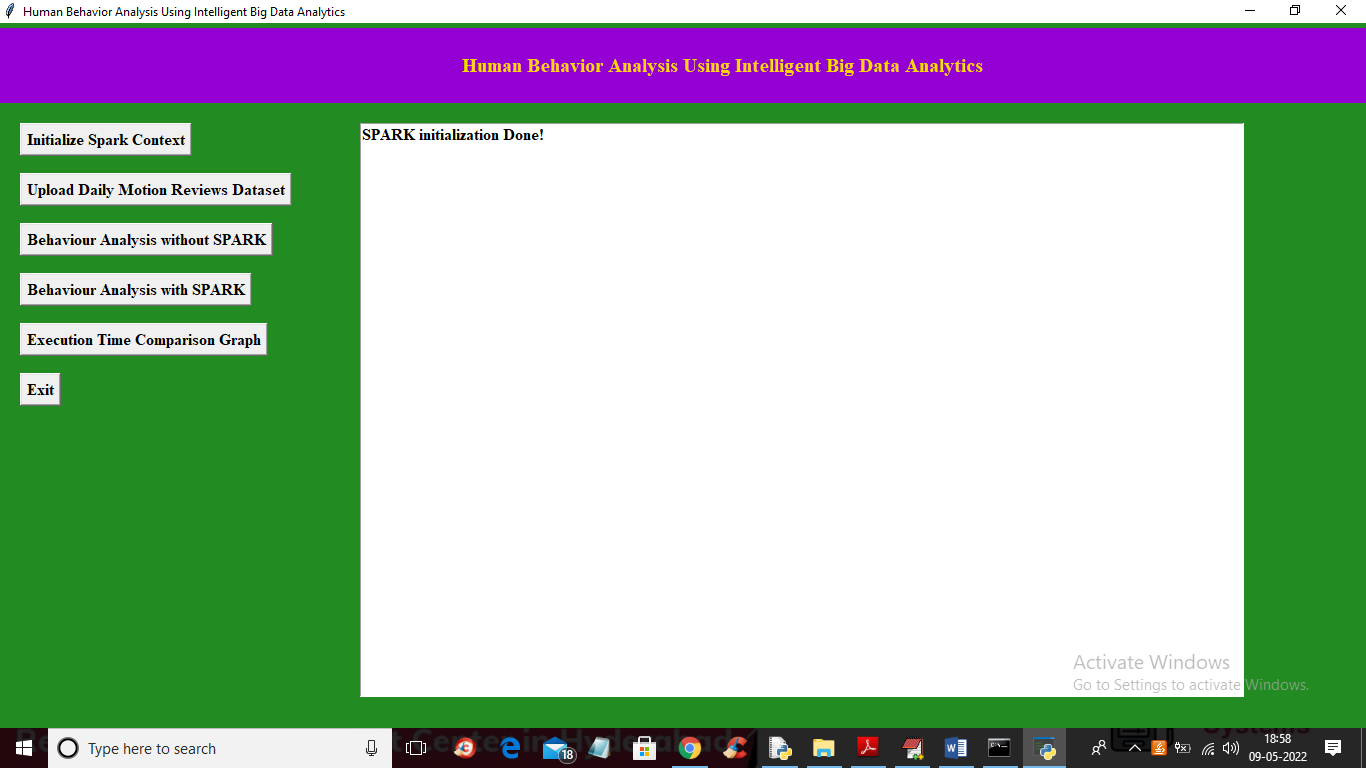
To run project double click on ‘run.bat’ file to get below output



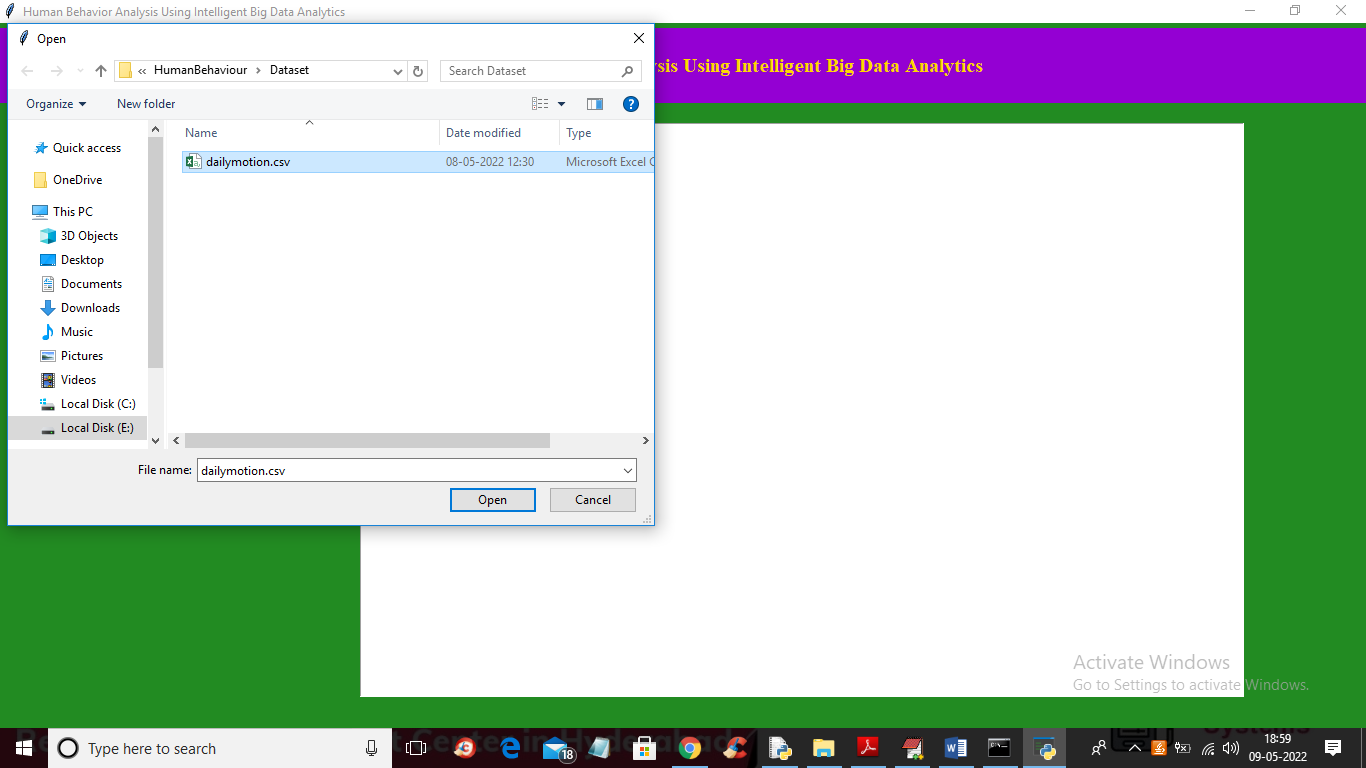
In above screen click on ‘Initialize Spark Context’ button to setup spark context and get below output after initialization



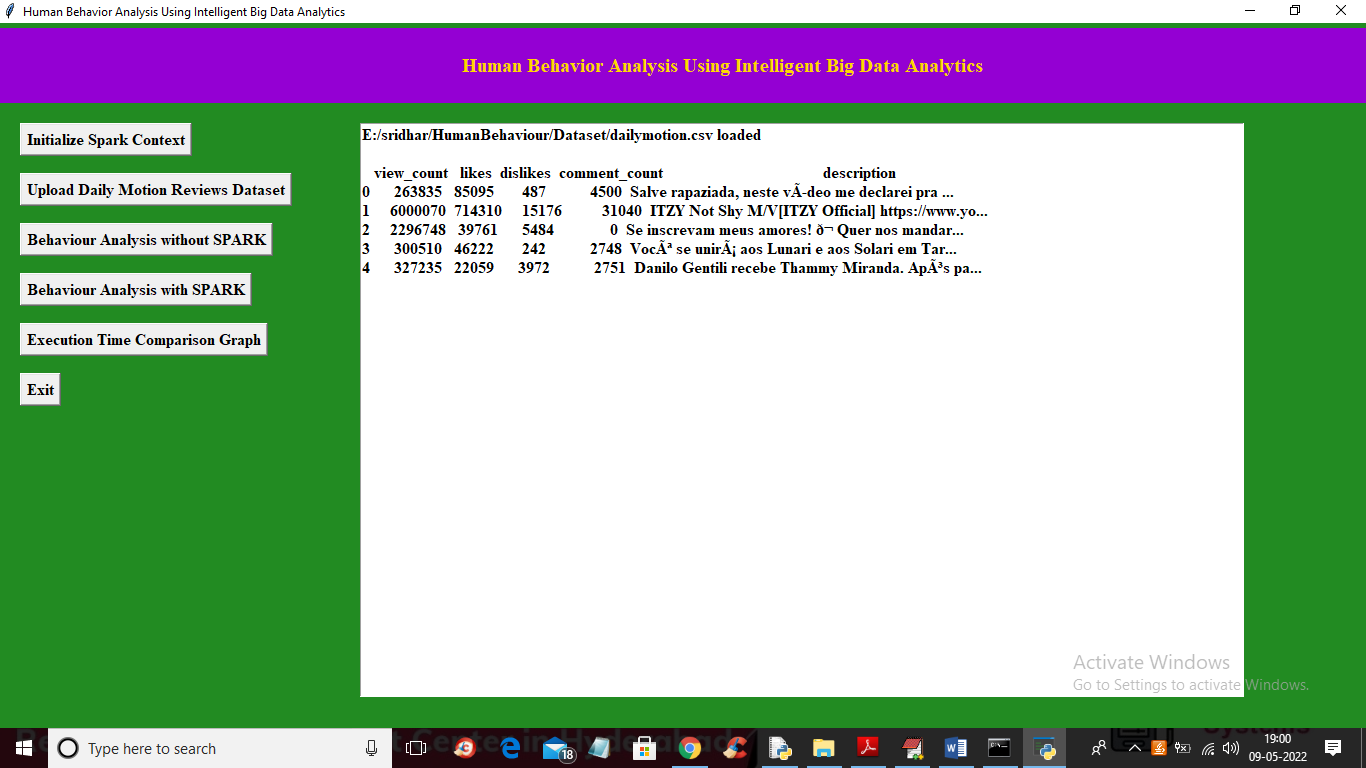
In above screen we can see SPARK object is getting initialized and after initialization will get below output



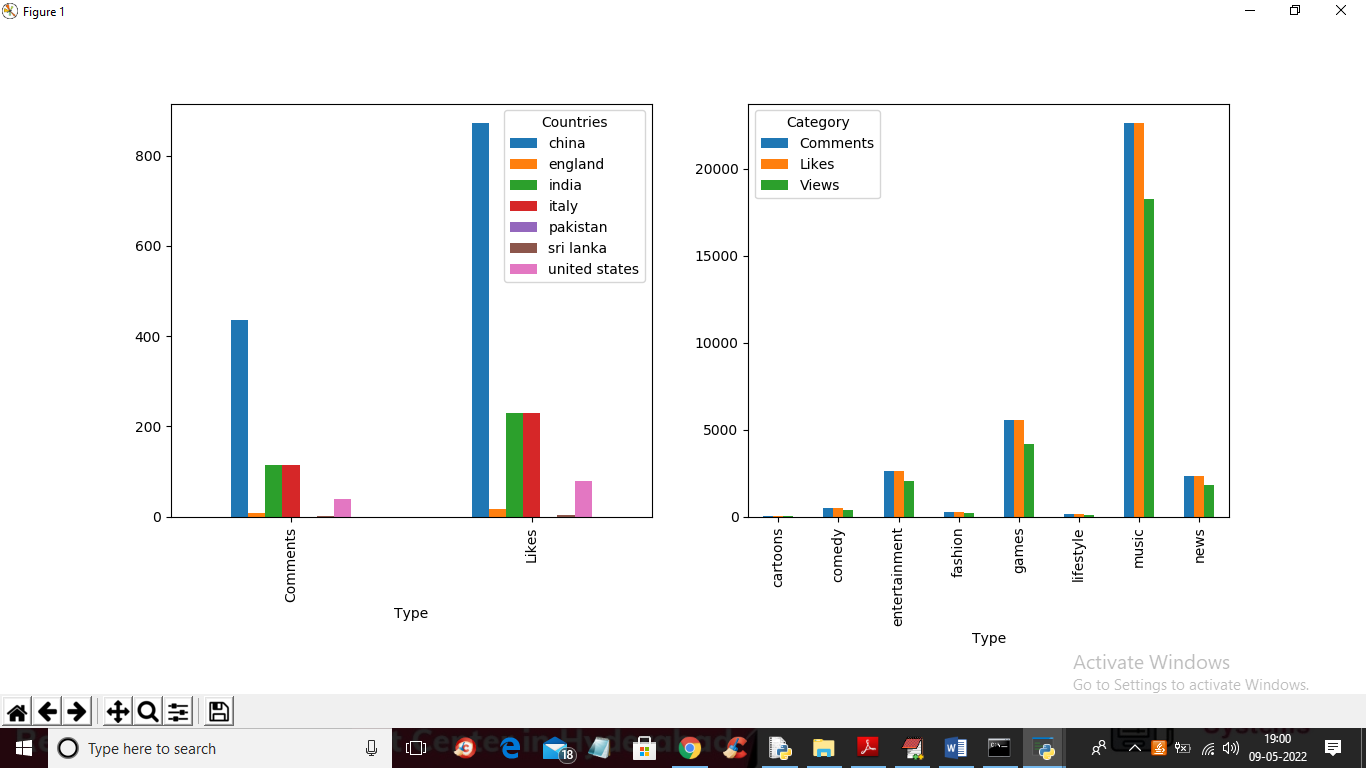
In above screen SPARK context is initialized and now click on ‘Upload Daily Motion Reviews Dataset’ button to upload dataset and get below output



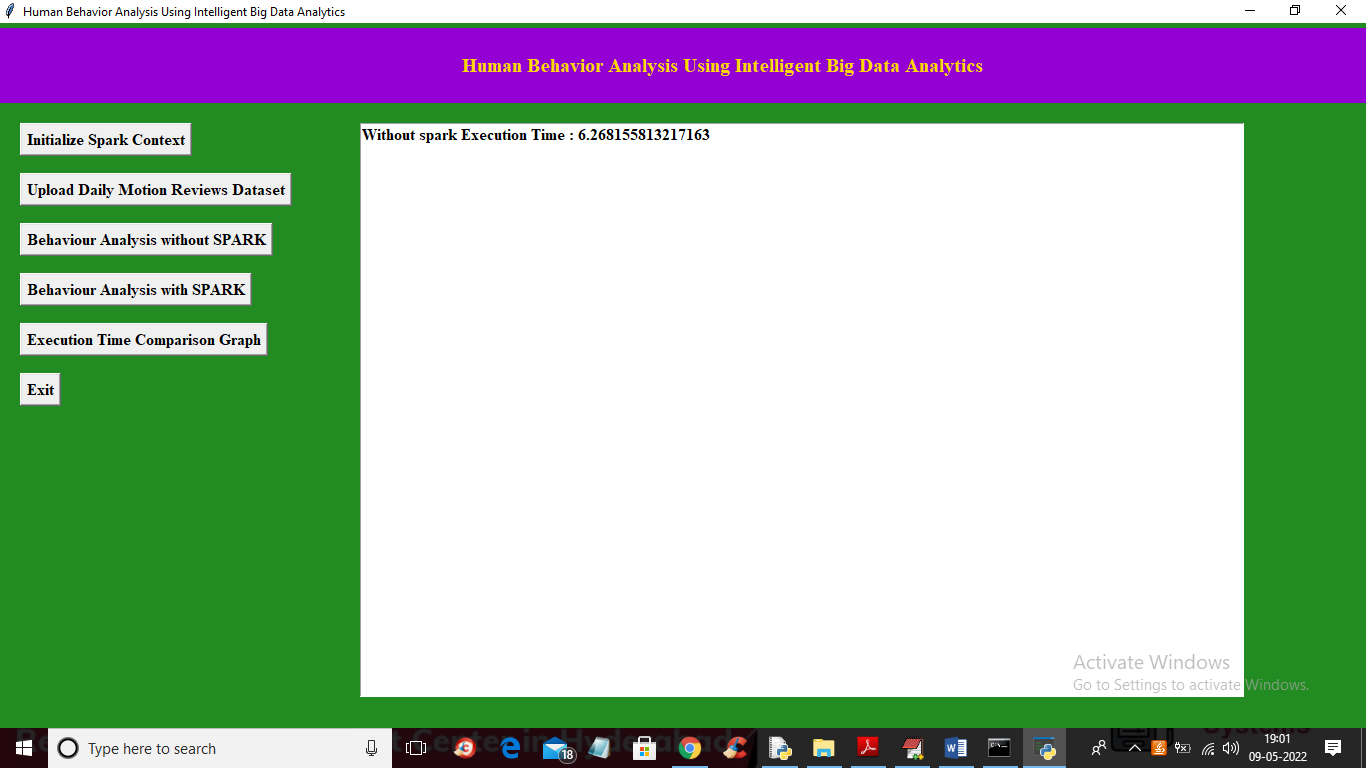
In above screen selecting and uploading dataset file and then click on ‘Open’ button to load dataset and get below output



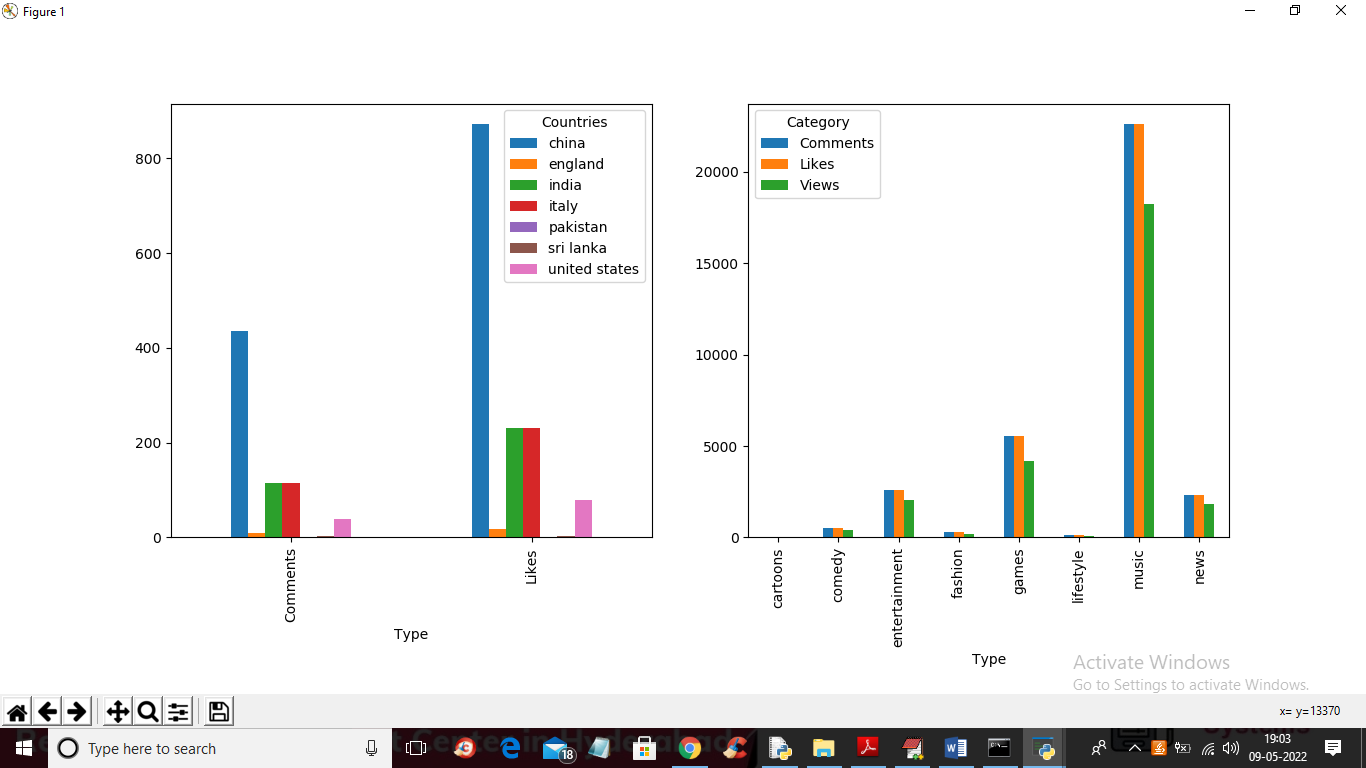
In above screen dataset loaded and now click on ‘Behaviour Analysis without Spark’ button to get below output



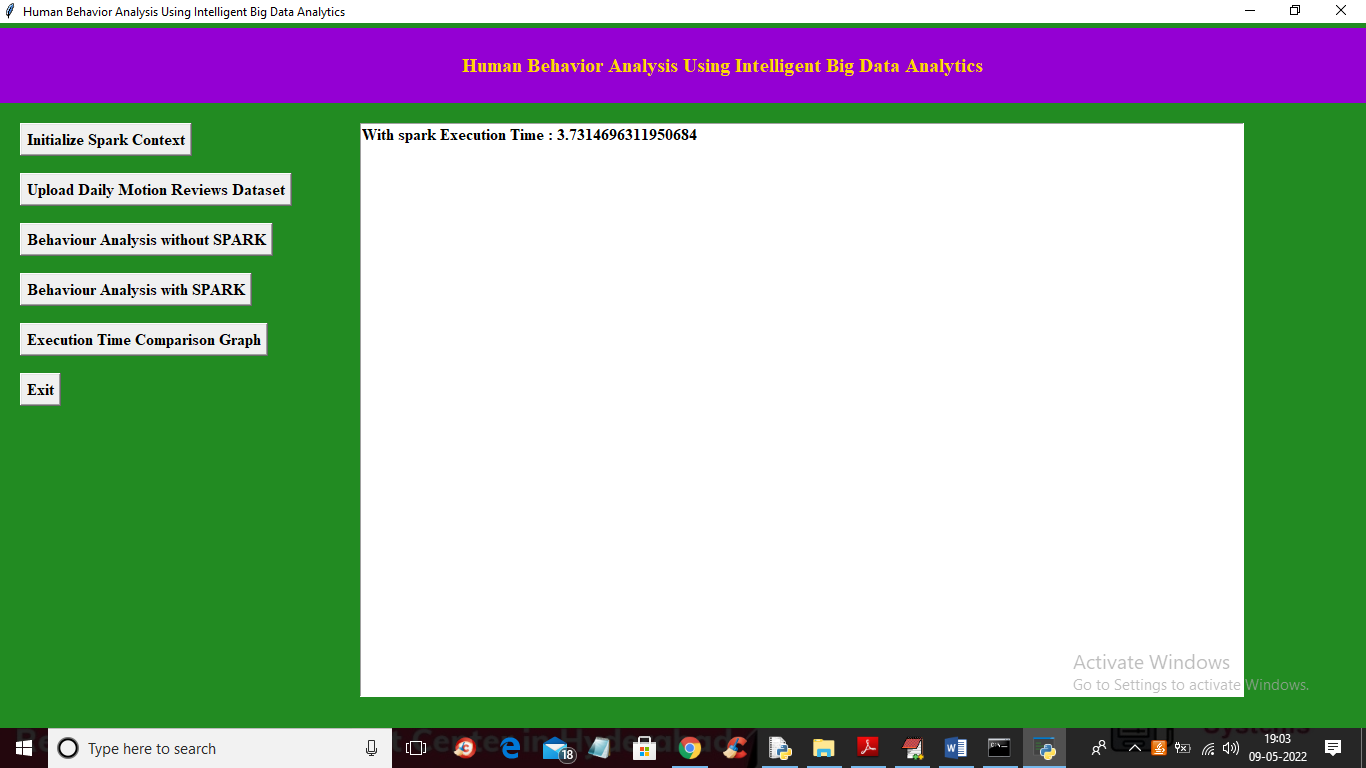
In above graph we have identify behaviour of persons like on which fashion or country they talk most with more LIKES and below screen showing execution time of WITHOUT spark processing



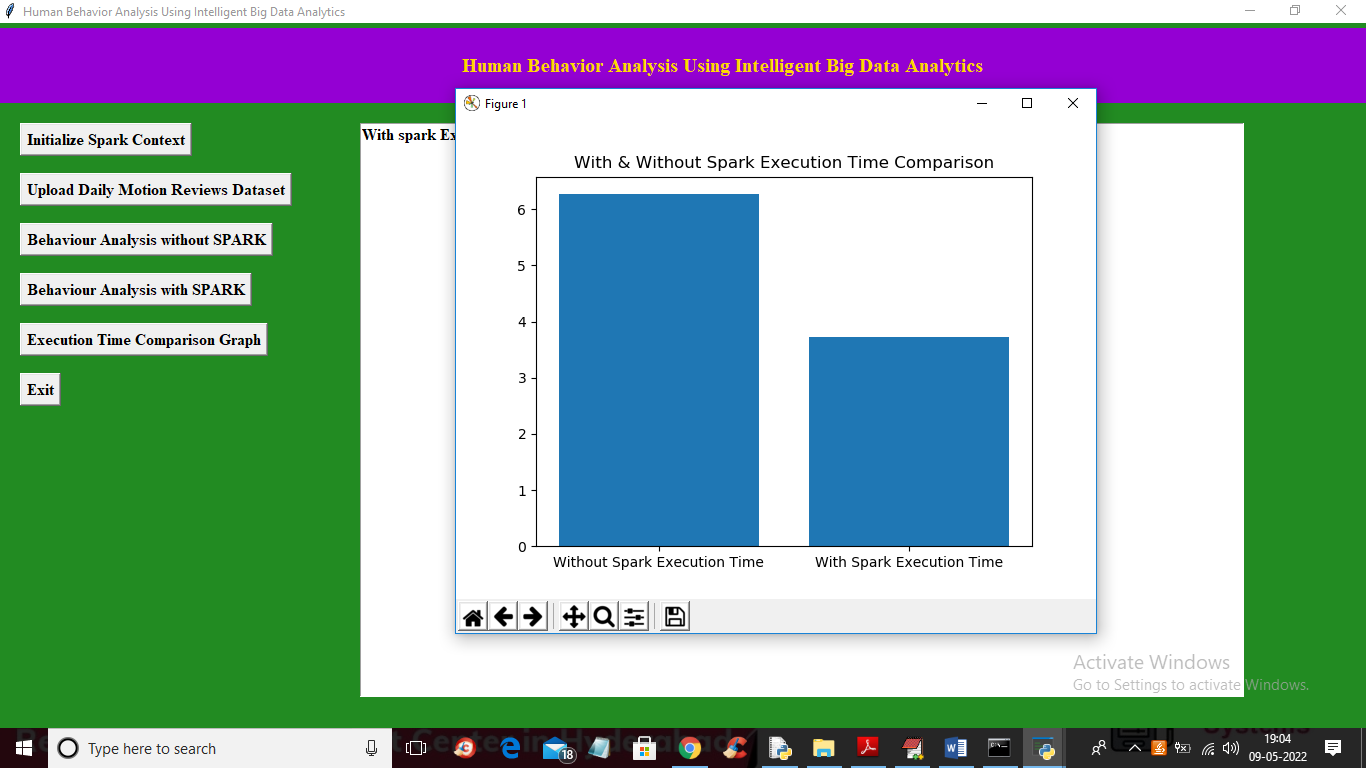
In above screen without spark processing it took 6.26 seconds and now click on ‘Behaviour Analysis with SPARK’ button to process same data using SPARK and get same output



In above screen with SPARK also we got same output but the difference is execution time and in below screen I am showing SPARK execution to process same data



In above screen SPARK took 3.73 seconds which is lesser than existing single thread processing and now click on ‘Execution Time Comparison Graph’ button to get below output



In above graph x-axis represents technique names and y-axis represents execution time and we can see SPARK processing took less execution time so it’s faster than traditional processing so BIG DATA processing with SPARK can be efficient