Work Portfolio

**Sachin Tiwari - BigData+Machine Learning+Analytics**

I am senior Big Data Analyst having 7+ years pf experience in data Analytics , machine learning, predictive modelling etc.

Below are tools and Technologies i have worked on so far

Analytical softwares :-

|  |  |  |  |
| --- | --- | --- | --- |
| AnyLogic | R |  |  |

Analytical languages/tools :-

|  |  |  |  |
| --- | --- | --- | --- |
| Hadoop | Spark | Scala | Hive |
| Pig | Kafka | Impala | Sqoop |
| Flume | System Dynamics | Fuzzy logic | Machine learning |
| Linear programming | Integer programming | Decision trees | Visulatiozation tools |
| Tableau | Looker |  |  |

Programming languages that i use :-

|  |  |  |  |
| --- | --- | --- | --- |
| Java | Scala | Python | VBA |
| JavaScript (Google sheets api) | SQL | R |  |

I have worked on complex analytical projects into my organisation.

i have worked on creating 300 node cluster for one of my projects in the organisation.

handling and analyzing data on this 300 node cluster helped me to learn and tackle real time issues.

I look forward to work with you.

Project Details

1. **Live Stream Analysis- FRM**

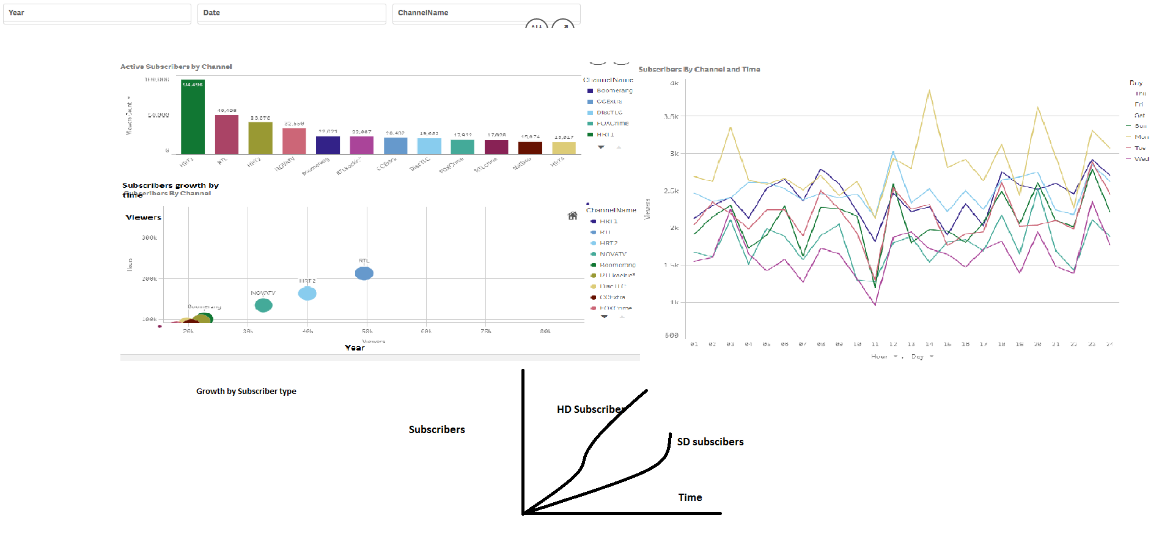
Apache MiNiFi agent will push the VCAS event logs (Filtered by Asset Site ID). MiNiFi Agent will be installed at retailer’s end via rpm.

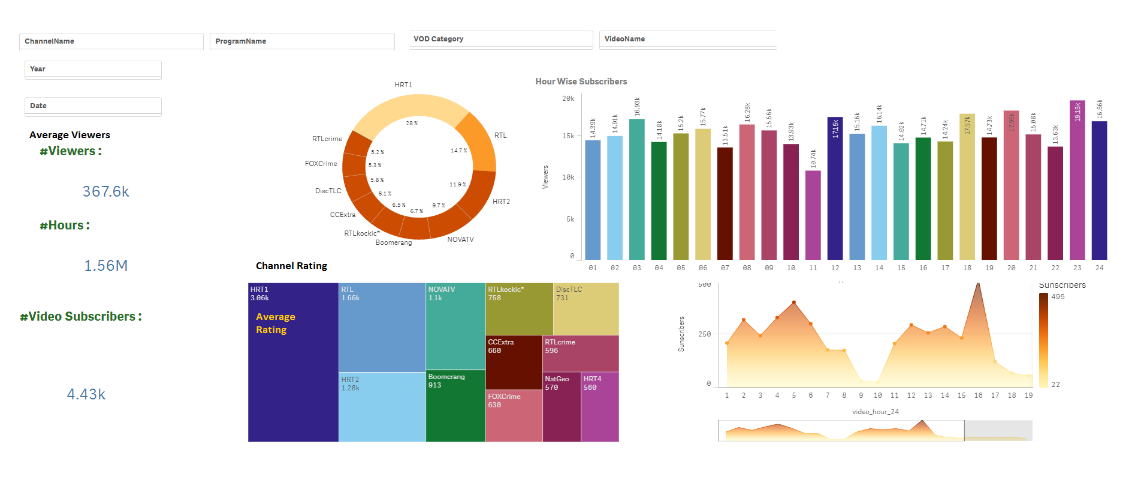
The event logs from MiNiFi agent will be push to AWS S3 over HTTPS/SSL (Added security to be discussed). All the retailers will be able to push the data via a common entry point AWS API Gateway to S3 buckets.

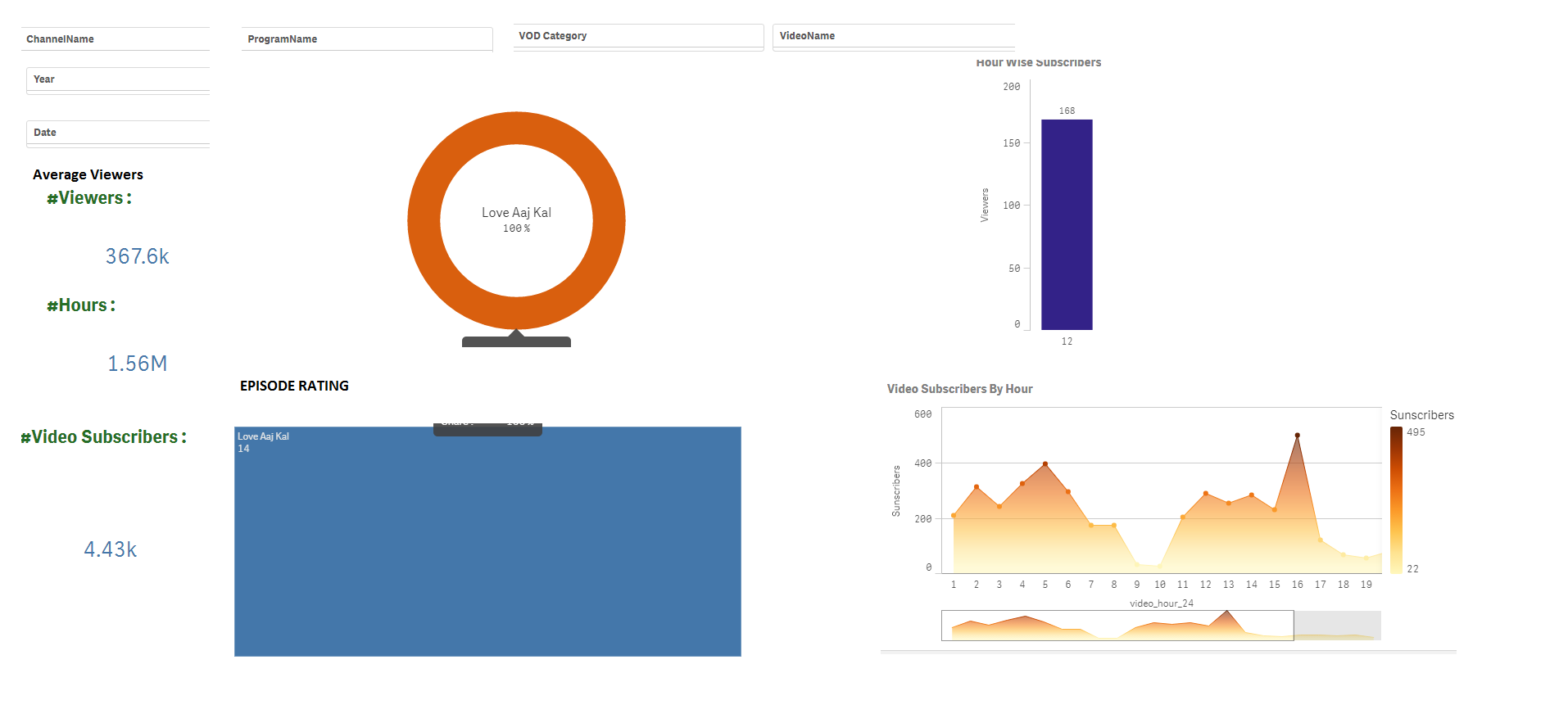
Apache NiFi will pull the data from the individual S3 bucket via (AWS SQS) and push in HDFS (AWS EMR). Each of the tenants will have their own data lake storage and summarization.The data will be loaded to the partitioned Hive table in AVRO format and the Spark job will perform the summarization.The Spark job will push the summarized data to the tenant specific Snowflake (Cloud based DWH). Tableau will provide reports to the end users using the data from the Snowflake DWH.

Technology stack used :-

|  |  |  |  |
| --- | --- | --- | --- |
| **Nifi** | **MiNifi** | **AWS** | **Hadoop** |
| **Spark** | **Tableau** | **Scala** |  |

1. ScreenShot-1:- Vod Actie Suscriber Bar and Line Graph
2. ScreenShot-2 :- Channel rating and hourly Subscriber Graph





1. ScreenShot-3 :- hourWise Video Subscriber Graph
2. **Tweet Sentiment Analysis**

This project predict user behavior/sentiment for specific incident happen around.

It was a real time Sentiment analysis. This project used Spark Mlib Library to implement machine learning. Using machine learning we were able to convert twitter raw feed to some meaningful data and rank the data on bases of keywords in various categories. Finally data connected to tableau to visualize and generate charts and graph depicting user behaviors for the specific incident happened real time and trended on twitter.

Below are the steps project was developed in

• Ingestion of Twitter data in hadoop using flume Ingestion from Twitter Stream.

• Hive and Pig to cleaning the data and storing into Hadoop. (Intermediate storage/staging/Landing)

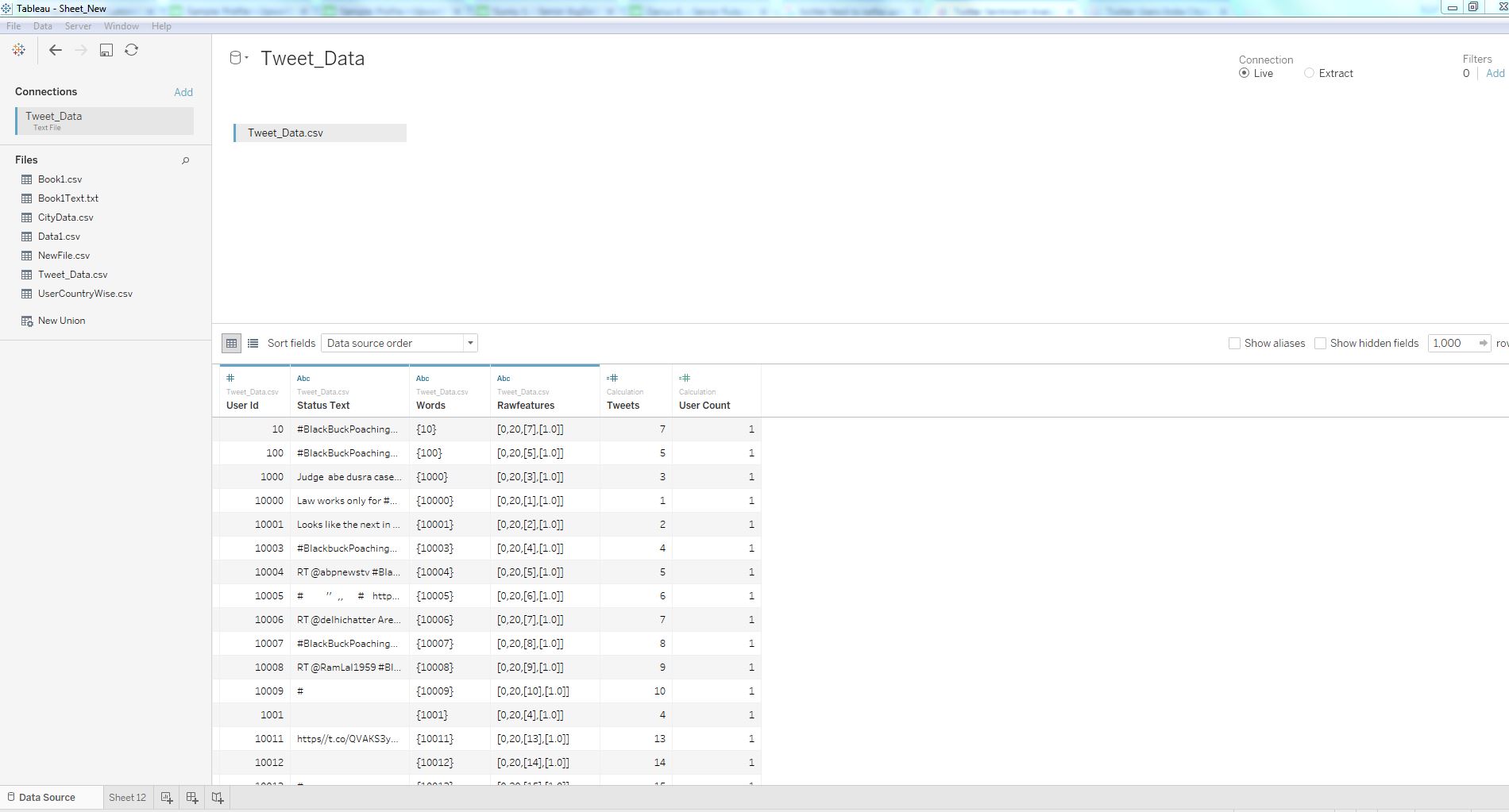
• Spark Mlib(Machine Learning) to process and load the data from hive tables to PostgreSql.

• BI Reports (Sentiment Analysis) will fetch the data from PostgreSql to the end users.

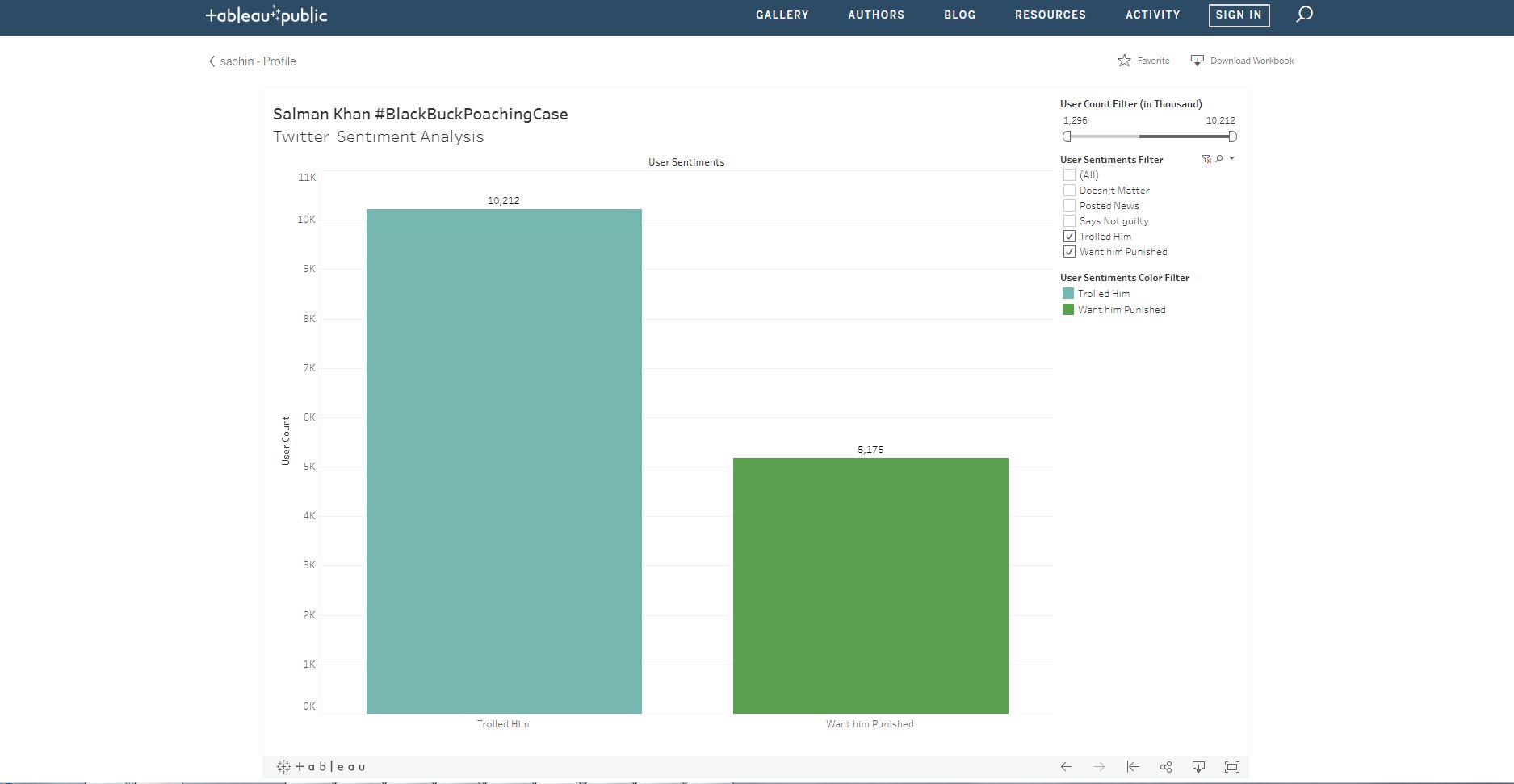
Final Tableau Reports :-

https://public.tableau.com/profile/sachin5164#!/vizhome/TwitterSentimentAnalysis\_6/TwitterSentimentAnalysis

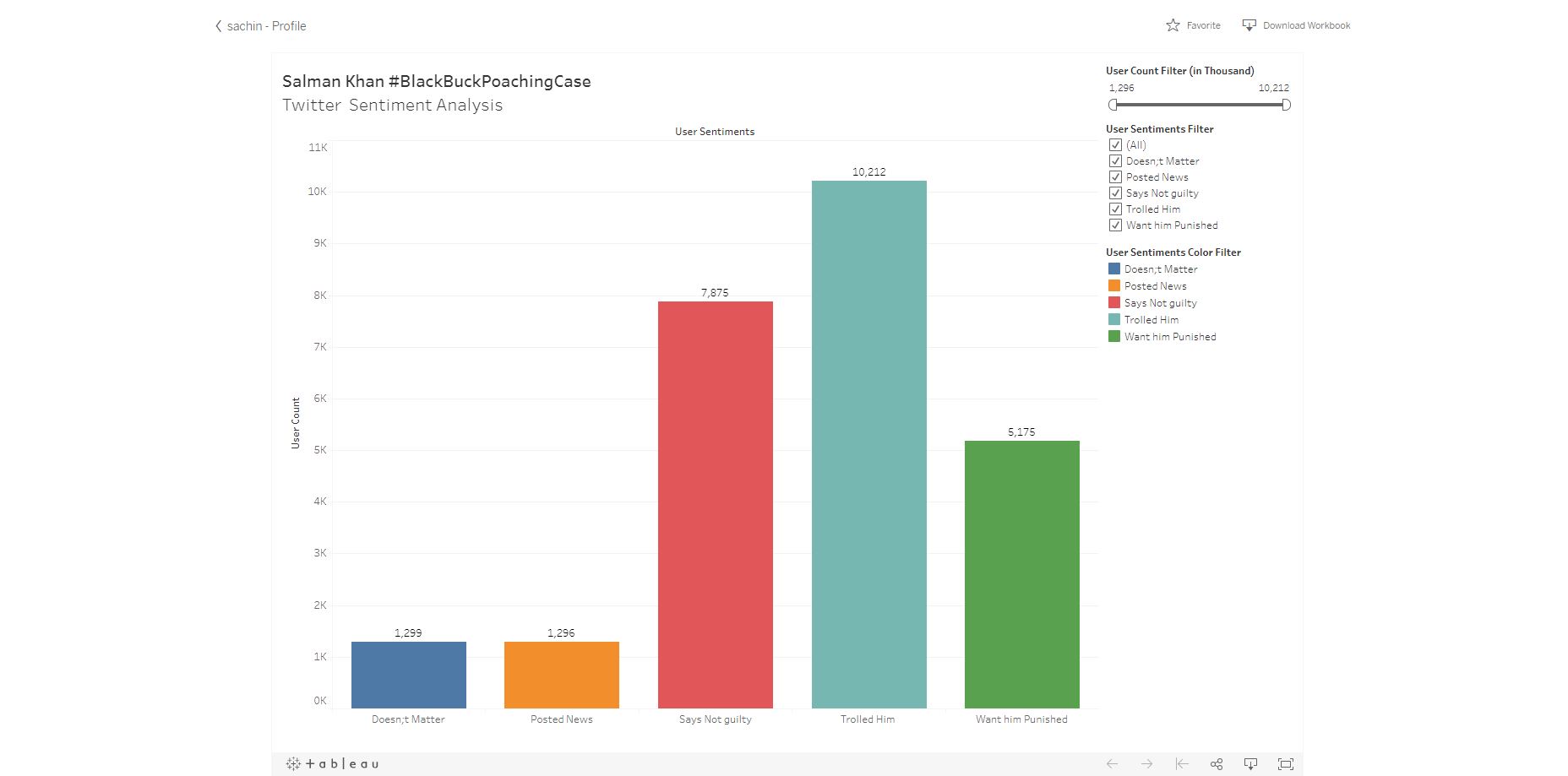
https://public.tableau.com/profile/sachin5164#!/vizhome/IPL\_40/Dashboard2

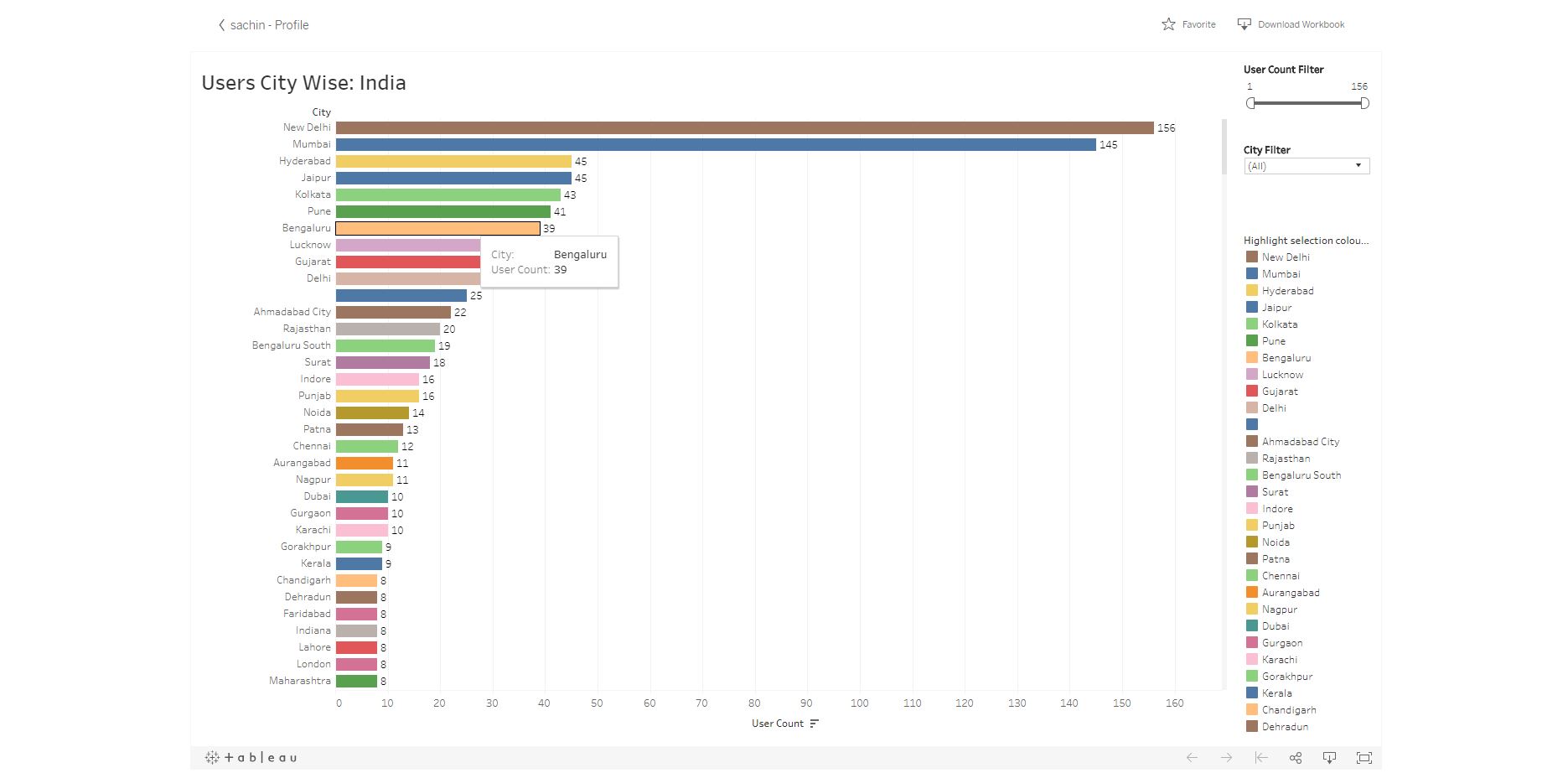
A. Screen Shot-1 Formatted Tweet Data in Tableau

B. ScreenShot-2 :- Tweet Predictive Analysis bar Graph in Tableau



1. ScreenShot-3 :- Tweet Predctive Analysis Bar graph in Tableau



1. ScreenShot-4:- :- Tweet Predctive Analysis Bar graph in Tableau