Alexandra Norman, Lauren Darr, Ramesh Simhambhatla , John Heinen

[[1]](#footnote-1)

Tableau Performance with Big Data Sets(Summer 2017)

# **INTRODUCTION**

**T**he intent of this paper is to provide the work evidence and conclusions for group project on Tableau Performance with Big Data Sets, undertaken as part of MSDS Database Management study

# **scope**

The scope of this project is to test the performance of Tableau with large data sets (or Big Data) from Social Media or with publicly available data sets.

**Notes: We have chosen US Census data on people and housing for 2011-2015. This is a large data set with nearly 300 variables for people, millions of records, over 100 variables on housing with millions of records.**

# **assumptions**

. Using methods learned so far in the MSDS program, we would like to learn how to access social media or publicly available data sets through API's using R or Python code. If time permits, we would like to look into using a NoSQL database to house this data based on a database schema that we develop. We will use a SQL database if during analysis we find that a NoSQL database is not possible in the time we have to complete the project. We would like to test the performance of our database and also test the use of Tableau on our database. We found that this project would apply to the work that several of our team members do using Tableau.

**Notes: We have completed a proof of concept using R and Python to load a MySQL database with a small subset of data. 36 variables and 10,000 records.**

**We have also tested that we are able to download Tableau and Tableau can connect to the database that we have created.**

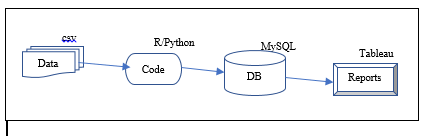
# **project plan**

A detailed step by step project plan was developed to estimate the tasks required and completed the project on time.

1. *Planning: Identified tools and work required, and work distribution among project members*
2. *Analysis: Perform feasibility study with required tools and experience; develop high level architecture*
3. *Design: Design database schema, and code*
4. *Implementation: Implement the designs and execute*
5. *Conclude: Complete the project and the paper with details and findings*

# **architecture**

A very high level diagram illustrate the workflow of the data import, save the data to MySQL database, and then test the Tableau report performance:



# **implementation**

Implementation of the design include 4 steps:

1. Develop database schema
2. Import and Save the data
3. Render data to Tableau
4. Test the performance

# **limitation**

<< update limitations and deviations from original scope due to whatever reasons >>

# **Conclusion**

## << update results, findings >>

**Appendix**

Appendixes, if needed, appear before the acknowledgment.

**Acknowledgment**

<< Acknowledge Prof, TA and peers & family/friends

Special Thanks to Professor Sohail Rafiqi, TA Nibhrat Lohia for the guidance and feedback etc >>.

**References**

1. *Reference 1*
2. Reference 2

1. TBD references, notes etc.

   Tbd references, notes etc [↑](#footnote-ref-1)