**Final Project: Responsibilities**

**Author’s Name:** Shivani Sharma

**Course Number:** ALY6110

**Course Title:** Data Management and Big Data

**Academic Term:** Fall 2019 CPS Analytics

**Instructor’s Name:** Mr. Daya Rudhramoorthi

**Assignment Completion Date:** 04-26-2020



As a part of our final project our team has been assigned clustering research area which we have decided to implement on the data set of a famous website “BlueBikes.com”. We have been provided with the data where riders rides, duration of rides, distance covered in the rides, popularity of the stations etc. We have data for each month. My roles and responsibilities include the “Data Visualization” and “Deployment” part of the project. I have decided to use Tableau majorly for preparing the dashboards. I might be using RShiny if required as we move forward in the project. Initially I have been analyzing the data set for finding out the patterns and correlation between parameters using Tableau. Based on the results I have decided to do analysis on clusters. Once the analysis regarding the correlation and clustering is done on initial data set I will provide the feedback to the team member who is going to write the algorithm logic for clustering and then once model is ready again I will pick up the dashboard creation on the refined dataset again for final representation of the results of the clustering that we will obtain.

I am also responsible for doing the data migration and deployment part of the project. So, the Sqoop import of the csv files into the Hadoop environment, table creation, and loading the data into Hive/ Impala whatever will meet our requirements. Then providing the details of the database to the developer of the algorithm. And finally reading the results of the output obtained. From the output we have decided to create the dashboard reports again and then providing the results to the team member who is going to write the report part.

**References**

1. <https://www.bluebikes.com/system-data>