Big Data Analytics Application Project Task List

Project Name	*Measuring Community Friendliness of Programming Languages*
Team Members	1) Samarth Tambad
	2) Rohit Nandwani
	3)

Step	Task	Who	Start Date	End Date	Comments
Data Planning Stage					
1	**Identify data sources**	***All***	Mar 25	Apr 3	
2	**Plan where data will reside**	***All***	Mar 25	Apr 3	*•How much total data will be stored at any given time?* *•Will the data fit in your Quickstart VM's HDFS, or do you need to use Dumbo? or the Cloud?*
3	**Plan remediation action(s)**	***All***	Mar 25	Apr 10	*•What action will your application take in response to the insight discovered? Will it be automatic or does it require user action?*
Data Cleaning and Profiling Stage (*Each data source is owned by and processed by one team member - steps 4, 5, 6)*					
Data Source 1 Processing					
4	**Write code to ingest data source 1 (if just a copy operation, please provide the command(s))**	Samarth	6 April	_	*•In this step, you'll read the data from the source and write it or copy it into HDFS*
5	**Write code to clean/format (ETL) data source 1**	Samarth	_	_	
6	**Write code to profile data source 1**	Samarth	_	13 April	*•This is to characterize the data and the range of values in each column •You might notice unexpected values in a column - you may decide to normalize the values (e.g. Street vs. St. vs. street) in the ETL stage •Find min, max, and averages •Find min and max length of text fields*
Data Source 2 Processing					
4	**Write code to ingest data source 2 (if just a copy operation, please provide the command(s))**	Rohit	6 April	_	
5	**Write code to clean/format (ETL) data source 2**	Rohit	_	_	
6	**Write code to profile data source 2**	Rohit	_	13 April	
Data Source 3 Processing					
4	**Write code to ingest data source 3 (if just a copy operation, please provide the command(s))**	NA	NA	NA	
5	**Write code to clean/format (ETL) data source 3**	NA	NA	NA	
6	**Write code to profile data source 3**	NA	NA	NA	
Data Source *n* Processing					
4	**Write code to ingest data source n (if just a copy operation, please provide the command(s))**	NA	NA	NA	

5	**Write code to clean/format (ETL) data source *n***	NA	NA	NA	
6	**Write code to profile data source n**	NA	NA	NA	
Analytic and Application Development Stage					
Develop UI / Visualization	**Design the UI / Visualization**	Rohit	14 April	_	
	Code the UI / Visualization	Rohit	_	_	
	Test the front-end UI / Visualization	Rohit	_	16 April	
Develop Analytic	**Design the back-end analytic**	Samarth	14 April	_	
	Code the back-end analytic	Samarth	_	_	
	Test the back-end analytic code	Samarth	_	_	
	Analyze *results* produced by analytic	Samarth	_	_	• Test your hypothesis - does it hold? • Are the results what you expected? • Iterate to improve results, and/or to better understand results
	Iterate on the analytic	Samarth	_	20 April	*• Iterate to improve results, and/or to better understand results • Do the results make sense? Are the results actionable?*
Develop Application	**Design the applicaton**	Rohit	17 April	_	*• How does the application talk to the analytic/analytic results? • How do the results get to the UI/Visualization?*
	Code the applicaton	Rohit	_	_	
	Test the applicaton	Rohit	_	20 April	
Final application code due		***All***	**-**	***May. 3, 2020***	