

December 7, 2024

1 Citations

In completing this project, several resources were utilized, including datasets, tools, and documentation. The following are the key references that contributed to the successful completion of this project:

1. **Dataset**: - Jha, S. (2021). **BigBasket Entire Product List (28K datapoints)**. Kaggle. <https://www.kaggle.com/datasets/surajjha101/bigbasket-entire-product-list-28k-datapoints>.
2. **PySpark Documentation**: - Apache Spark. **PySpark Documentation**. Apache, 2021. <https://spark.apache.org/docs/latest/api/python/>.
3. **Tableau for Data Visualization**: - Tableau Software. **Getting Started with Tableau**. Tableau, 2021. <https://www.tableau.com/learn/training>.
4. **MapReduce and Hadoop Documentation**: - Apache Hadoop. **MapReduce Overview**. Apache, 2021. [https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html](https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html).
5. **GitHub Repository**: - Prameela2511. **Big Data E-commerce Project**. GitHub Repository. [https://github.com/Prameela2511/4517_Bigdata_E-commerce_project.git](https://github.com/Prameela2511/4517_Bigdata_E-commerce_project.git).

2 GitHub Repository

All the source code, documentation, and project artifacts are available in the GitHub repository. You can access and review the entire project at the following link:

https://github.com/Prameela2511/4517_Bigdata_E-commerce_project.git