MSDS442
Project – Phase 4

Submission:

Submit your Project-Phase 4 as a SINGLE ZIP file on Canvas by the due date.
Your submitted ZIP file must have the name:
Project_Phase_4_Your_LastName.zip

Deliverables:

Your ZIP file for the exercise submission must include the following:

- All source code that you wrote, compiled and built on your personal computer.
- Panopto video recording of a live run of your code on your personal development computer.

Requirements specification:

Utilize **Kafka**, **Java**, **Python**, **Facebook/Prophet**, **TensorFlow Time Series (TFTS)**, **StatsModel** and **Neo4j GDS** library to implement the following requirements based on the requirements specification and architecture document for the OnMart Superstore real-time data streaming application that you reviewed in Phase 1:

1. Create Customer-Friends-Product-Reviews graph database for OnMart

2. Use Google **PageRank**, and the graph database that you created above for customer reviews/ratings to find **influential reviewers**

3. Use **Betweenness Centrality** and **PageRank** algorithms to find which OnMart superstore super-connected and influential reviewers who can be used to identify products from the **Furniture department** that it could use in a cross-promotion with the newly introduced Samsung TV in the **Electronics department**

4. The day of the week that has the maximum number of purchases made

5. The month of the year that has the highest sales revenue

6. The weekly sales revenue forecast per zip-code

7. The daily number of purchases forecast per distribution center

8. The weekly number of product returns forecast per warehouse

9. Whether the number of delayed deliveries per zip-code positively correlated to the number of returned purchases/products.

10. The busiest zip code (counting both deliveries and purchase returns)

11. The most selling product in every zip-code

12. Analyze whether the sales of the Electronics department and the Clothing department have similar seasonal patterns

13. Product purchases seasonality through the year; for example, Back-to-School, Christmas, Superbowl, Weather, etc.

14. Build time series analysis model for product X (specify product) in zip-code Y (specify zip code) to forecast daily sales revenue