

Data Management II – ETL and Smart Data Project Submission Template

Name of the Student:	<u>Jaime Paolo Vega Arellano</u>
Matriculation Number:	<u>11015424</u>
Dataset Considered for Project:	<u>ECommerce</u>

Note:

- All the 3 questions mentioned below must be answered by the dataset considered.

Question 1:

Is Star Schema or Snowflake schema or Fact Constellation Schema possible with the dataset? If yes, then which schema? If no, then drop a mail regarding the issue with reason.

Answer:

Snowflake schema

Question 2:

Which all quality dimensions are covered by the dataset? Mention the data quality dimensions for the relevant columns and also provide a summary of data profiling for the dataset?

Answer:

Consistency, Validity, Uniqueness, Precision and Conformity

Question 3:

Question that will be answered from the dataset using data analytics techniques?

Explain in 2-3 lines the question that will be answered and also specify under which data mining model the question will be classified.

Answer:

1. Customer Segmentation. The Marketing team needs a way of classifying current customers to target those young and actively engaged in purchasing more products. A clustering algorithm is considered in this scenario.
2. Product Recommendation. The operations department needs to add a feature to their website to remain competitive in the market and recommend products to their users. Collaborative Filtering based on items is considered for this task.