

Template - CCS _CSPC



Camarines Sur Polytechnic Colleges

Document Details

Submission ID

trn:oid::23367:93802505

Submission Date

May 1, 2025, 1:51 PM GMT+8

Download Date

May 1, 2025, 1:52 PM GMT+8

File Name

Template - CCS _CSPC

File Size

2.6 KB

2 Pages





350 Words

2,288 Characters




11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

-  **4** Not Cited or Quoted 11%
Matches with neither in-text citation nor quotation marks
-  **0** Missing Quotations 0%
Matches that are still very similar to source material
-  **0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation
-  **0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 0%  Internet sources
- 0%  Publications
- 11%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 4** Not Cited or Quoted 11%
 Matches with neither in-text citation nor quotation marks
- 0** Missing Quotations 0%
 Matches that are still very similar to source material
- 0** Missing Citation 0%
 Matches that have quotation marks, but no in-text citation
- 0** Cited and Quoted 0%
 Matches with in-text citation present, but no quotation marks

Top Sources

- 0% Internet sources
- 0% Publications
- 11% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

- 1 Submitted works

City University of Hong Kong on 2024-11-04 6%
- 2 Submitted works

Coventry University on 2024-08-01 3%
- 3 Submitted works

Richfield Graduate Institute of Technology on 2024-09-04 2%

NAME: SEAN XANDER B. AQUINO, BSCS - 2A

The UCI Machine Learning Repository provides the Online Retail Dataset for this program. The dataset includes transaction data of a UK-based online retailer between December 2010 and December 2011. The dataset tracks purchase data of customers, mostly from wholesale businesses, on their product range.

Key Features:

InvoiceDate: Timestamp of purchase.

Quantity: Number of units purchased per product,

UnitPrice: Price of each unit, which is measured in GBP.

CustomerID: Unique customer identifier.

Description: Product name.

Country: Country of the customer.

A set of data-preparation procedures was needed before conducting clustering analysis. The data cleaning started by eliminating everything that lacked CustomerID values, because this step maintained a clear link between all transactions and their respective customers. Transactions containing negative values in Quantity and UnitPrice fields were removed, because they represented either returns or processing errors. Additional filtering of the dataset maintained only transactions involving United Kingdom customers, in order to conduct a consistent market analysis. TotalPrice served as a new calculation that computed the total transaction costs by multiplying Quantity with UnitPrice. Aggregation of customer-level features completed to produce three summarized behavior metrics that included TotalQuantity (items purchased), TotalSpent (transaction sum), and ProductVariety (unique items).

The analyzed data underwent KMeans Clustering analysis, as an unsupervised learning method to organize similar purchasing behavior customer groups.

The application of StandardScaler accomplished feature scaling.

Visual inspection will be possible through 2D representation, after applying Principal Component Analysis (PCA) to the data.

The Silhouette Score identified the right cluster number as $k = 3$.

The customer segments, formed according to TotalSpent, ProductVariety, and PurchaseFrequency data, received these labels:

High-Value Customers

Highest spending, most frequent purchases, and high product diversity.

Moderate Customers

Average spending and frequency, loyal but not premium buyers.

Low-Value Customers

Infrequent, low-spending customers with limited product engagement.

Using PCA, the 2D scatter plot displays visible clusters, which have been named after customer segments. This visualization helps businesses:

Administrators should select VIP customers for membership in loyalty schemes.

Target moderate buyers with promotions.

Companies should understand the threat of customer turnover in their low-value market segments.