Database Design and Development – ISYS2038- Individual Assignment

Due: 12th June 2024

Class: Wednesday, 10:30 am

Nikunj Gupta – s4027333– (s4027333@student.rmit.edu.au)

Tutor: Dr. Yee Ling Boo

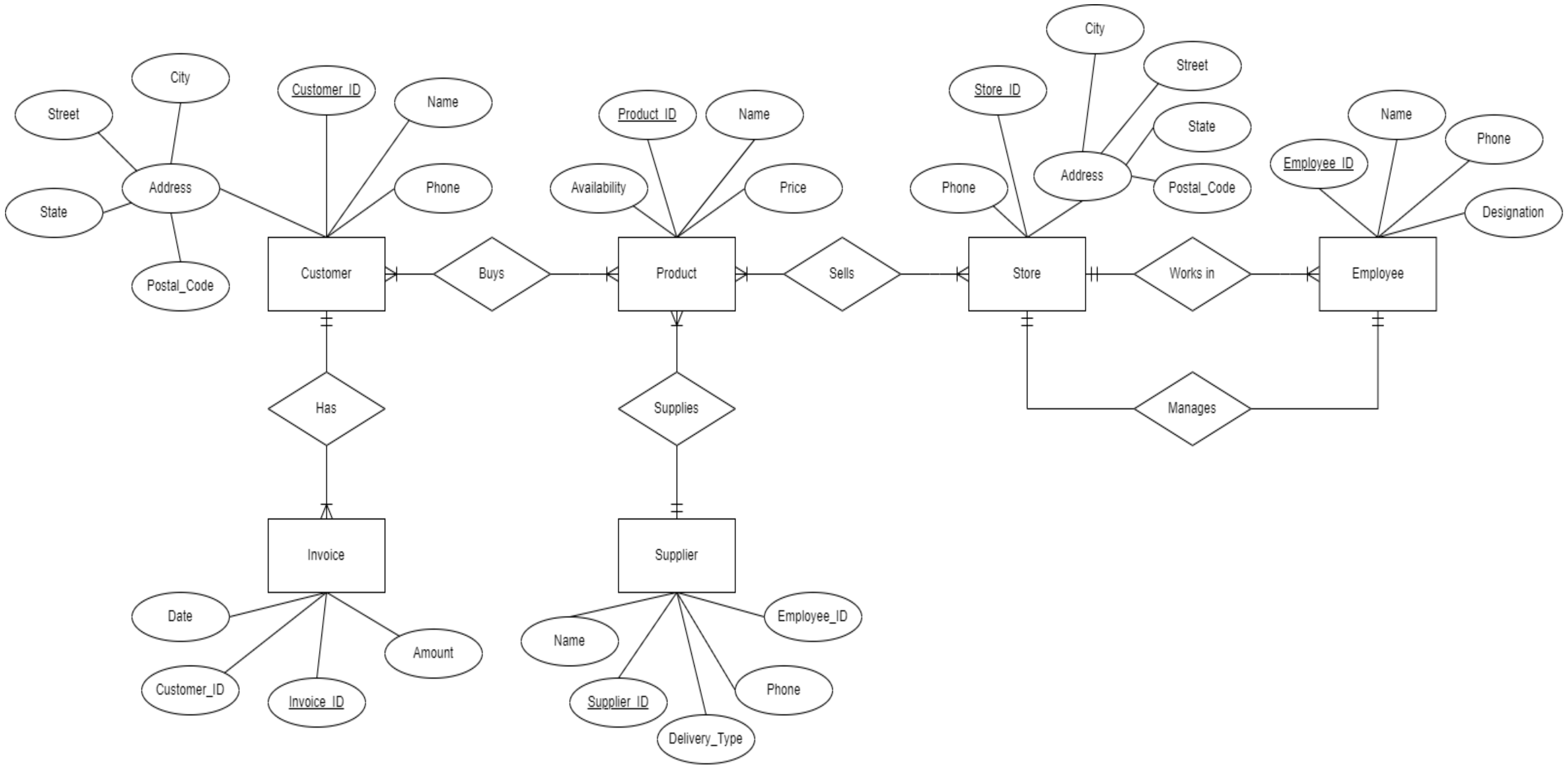
Word count = 2223

(Excluding tables, figures, Screen shots, References, Appendix, Table of Context)

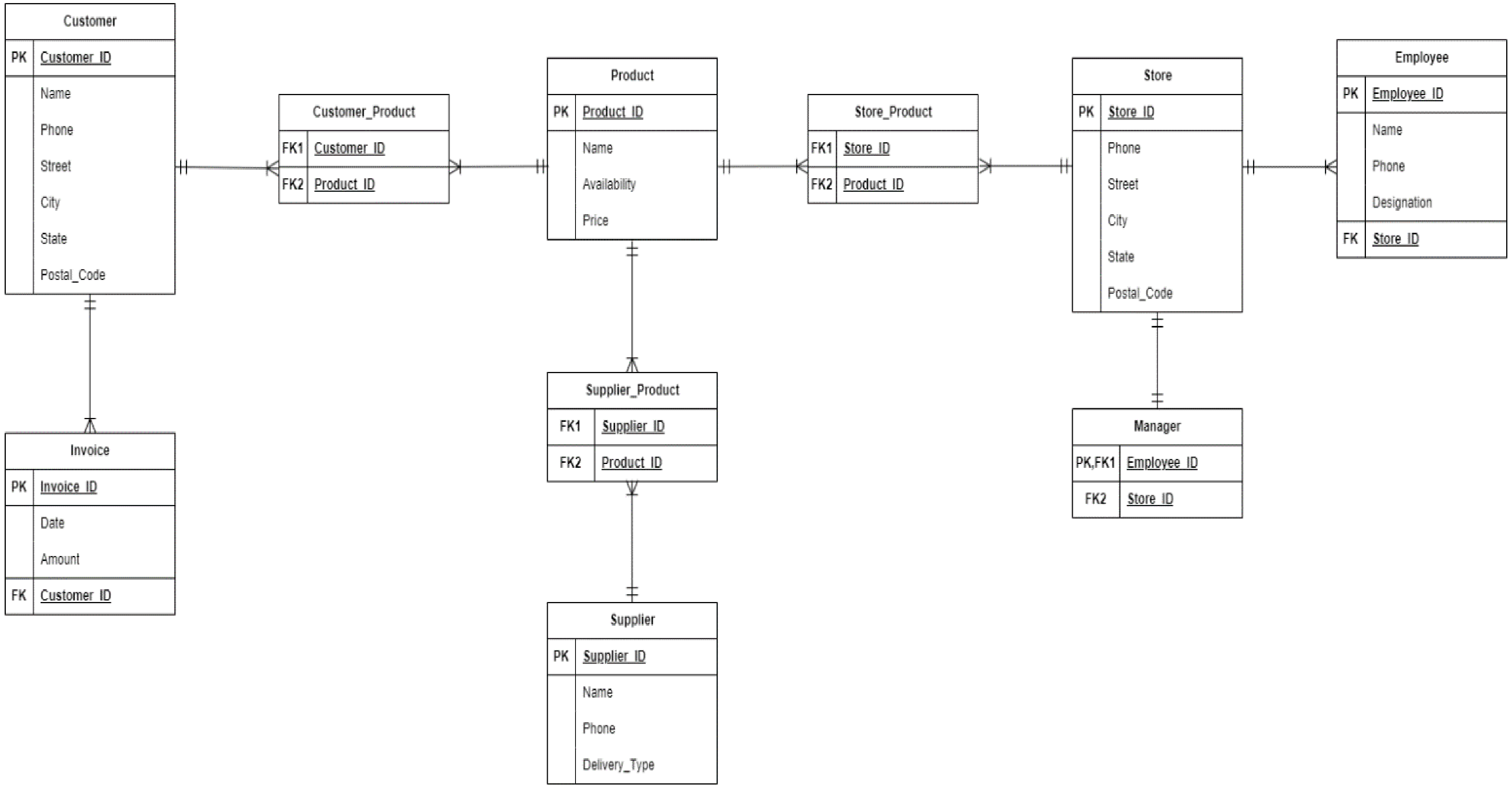
Contents

Question 1:

Entity Relationship Diagram:



Relational Model:



Question 2:

SQL Scripts:

1. Query with Calculation and Group By

```
SELECT p.Name AS Product_Name, SUM(I. Amount) AS Total_Sales

FROM Product p

JOIN Customer_Product cp ON p.Product_ID = cp.Product_ID

JOIN Invoice i ON cp.Customer_ID = i.Customer_ID

GROUP BY p.Name;
```

Sample Data:

Invoice Table:

Invoice_ID	Date	Amount	Customer_ID
1	2023-01-01	1500.00	1
2	2023-01-02	550.00	2
3	2023-01-03	200.00	1

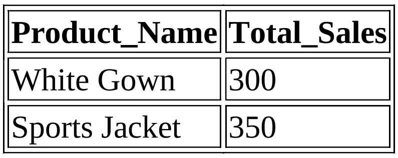
Customer_Product Table:

Customer_ID	Product_ID
1	1
2	2
1	2

Product Table:

Product_ID	Name	Availability	Price
1	White Gown	In Stock	900.00
2	Sports Jacket	In Stock	500.00

Potential Outcome:

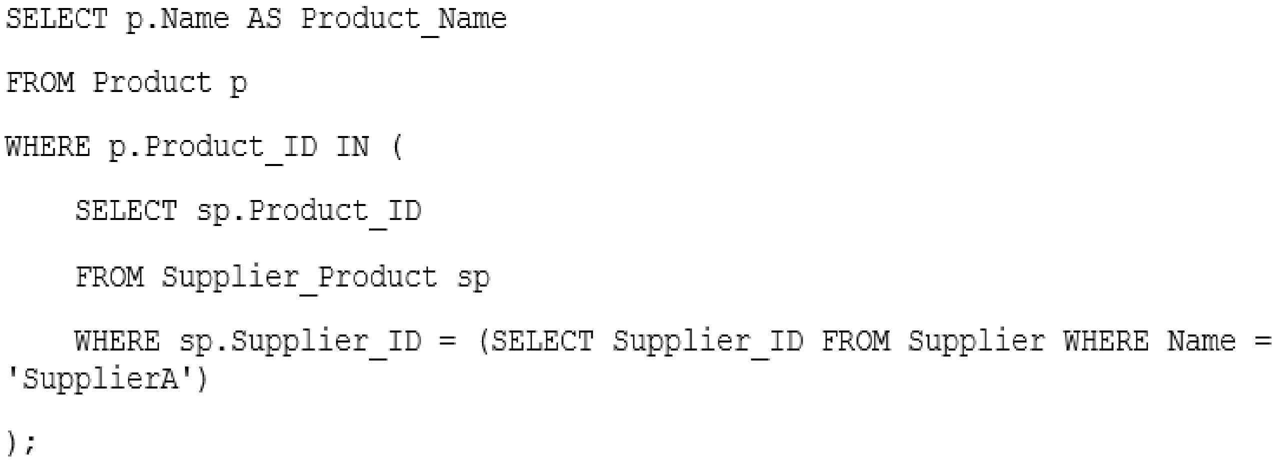


Product_Name	Total_Sales
White Gown	300
Sports Jacket	350

Business Purpose: Calculate the total sales amount for each product to understand which products generate the most revenue.

Explanation: Essentially, the SQL script retrieves data from three tables (Product, Customer_Product, and Invoice) in order to compute the aggregate sales for each individual product. The tables are joined based on the required keys (Product_ID and Customer_ID) and the results are grouped by product name to consolidate the sales amounts. The ultimate outcome displays the names of the products and their respective cumulative sales figures.

2. Query with Nested Query and Join



```
SELECT p.Name AS Product_Name

FROM Product p

WHERE p.Product_ID IN (

    SELECT sp.Product_ID

    FROM Supplier_Product sp

    WHERE sp.Supplier_ID = (SELECT Supplier_ID FROM Supplier WHERE Name =
'SupplierA')

);
```

Sample Data:

Supplier Table:

Supplier_ID	Name	Phone	Delivery_Type
1	SupplierA	123-456-789	Courier Van
2	SupplierB	987-654-321	Air

Supplier_Product Table:

Supplier_ID	Product_ID
1	1
1	2
2	1

Product Table:

Product_ID	Name	Availability	Price
1	Sports Jacket	In Stock	500.00
2	White Gown	In Stock	900.00

Potential Outcome:

Product_Name
Widget
Gadget

Business Purpose: List all products supplied by a specific supplier. This helps in managing supplier-product relationships and inventory.

Explanation:

Within the Stella Pty. Ltd. case study, the query can be utilised to ascertain all the products that have been provided by a particular supplier. Stella's firm entails overseeing numerous suppliers and a diverse array of products, making these enquiries crucial for comprehending supplier relationships and inventory management.

Tables in the database that are involved:

- The package contains comprehensive information about many items, encompassing their names and unique IDs.

- The Supplier_Product table establishes a relationship between suppliers and the products they provide, including the unique identifiers for each product (Product_ID) and provider (Supplier_ID).
- Supplier: Provides comprehensive information about suppliers, including their names and unique IDs.

The query facilitates effective management of supplier-product connections by enabling Stella to effortlessly retrieve information on the items supplied by a certain supplier. This, in turn, assists in inventory management and enhances the decision-making process.

3. Query with Scalar Function and Nested Query

```
DELIMITER //
CREATE FUNCTION GetMaxProductPrice(StoreID INT)
RETURNS DECIMAL(10, 2)
DETERMINISTIC
BEGIN
    DECLARE MaxPrice DECIMAL(10, 2);

    SELECT MAX(p.Price)
    INTO MaxPrice
    FROM Product p
    JOIN Store_Product sp ON p.Product_ID = sp.Product_ID
    WHERE sp.Store_ID = StoreID;

    RETURN MaxPrice;
END //
DELIMITER ;

SELECT s.Store_ID, s.City, GetMaxProductPrice(s.Store_ID) AS
Max_Product_Price FROM Store s;
```

Sample Data:

Product Table:

Product_ID	Name	Availability	Price
1	White Gown	In Stock	900.00
2	Sports Jacket	In Stock	500.00

Store Table:

Store_ID	Phone	Street	City	State	Postal_Code
1	043456789	1st Street	Richmond	VIC	3000
2	047654321	2nd Street	Brunswick	VIC	3145

Store_Product Table:

Store_ID	Product_ID
1	1
1	2
2	1

Potential Outcome:

Store_ID	City	Max_Product_Price
1	Richmond	500.00
2	Brunswick	900.00

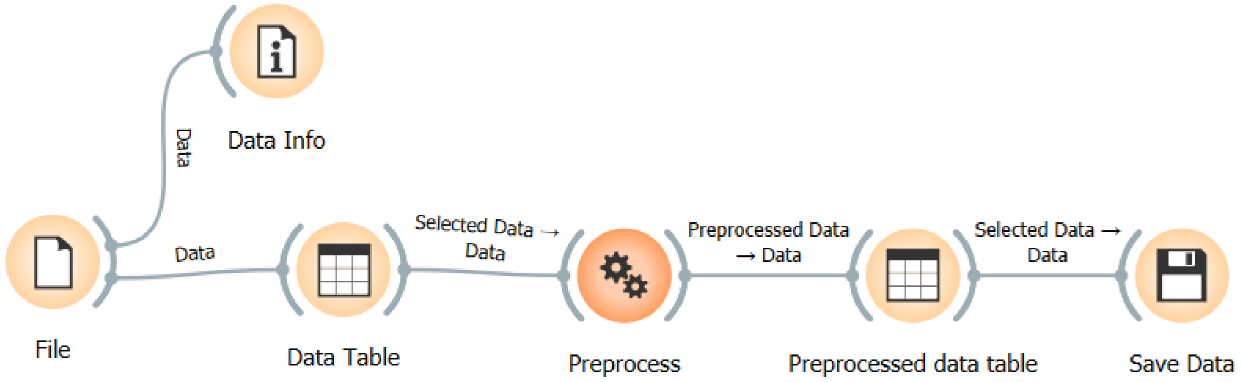
Business Purpose:

- Find the most expensive product sold by each store and its price.
- This function and query combination retrieves the maximum price of products sold by each store, helping store managers quickly identify the highest priced items in their inventory.

Explanation:

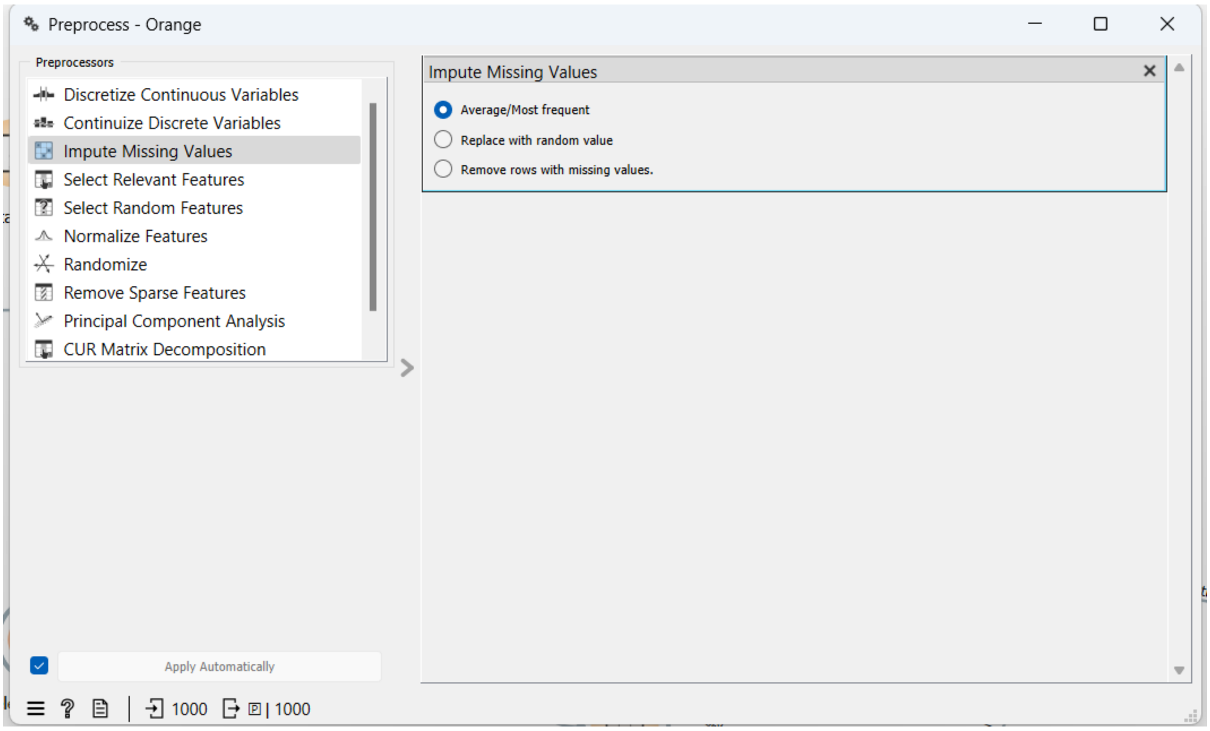
This function and query provided by Stella Pty. Ltd. can assist Stella Corlini and her employees in determining which retailer offers the most expensive merchandise. This can be advantageous for the purpose of managing inventory, formulating pricing strategies, and determining the stocking of high-value commodities. The inclusion of this feature in the database management system greatly enhances the efficiency of generating informative business reports for Stella Pty. Ltd., as it deals with a diverse array of products sourced from various suppliers.

Question 3: Data Analytics with Orange



Missing Values:

Within the Stella Pty. Ltd. dataset, there are 1,000 occasions where particular numbers are not present, and these missing values are indicated by the symbol "?". Although we can use excel files to predict data as planned, we need to initially employ orange to complete the missing amounts. To accomplish this, the Stella Pty. Ltd. dataset file is first imported into Orange, as shown below, and then the dataset is linked using the following set of linkages.



Impute Missing Values

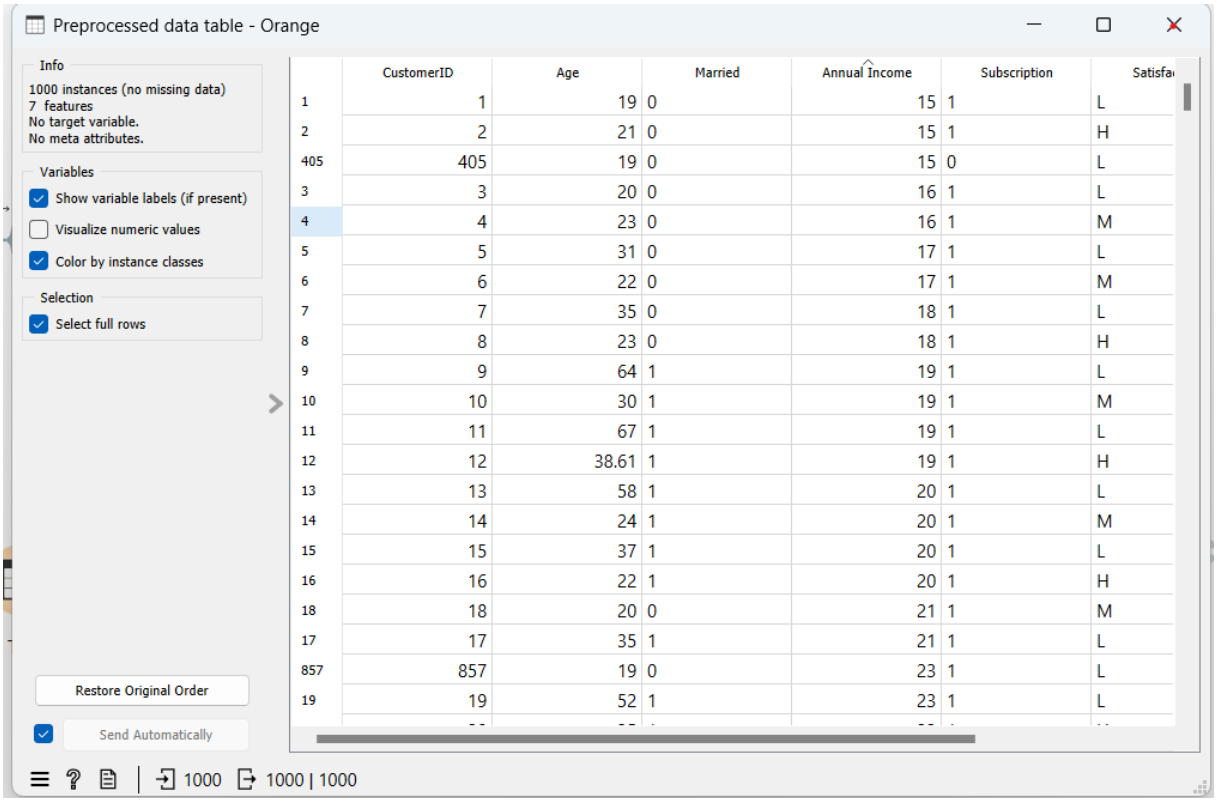
● Average/Most frequent

○ Replace with random value

○ Remove rows with missing values.

☑ Apply Automatically

The purpose is to fill in missing values in the dataset using the most commonly used method. This guarantees that the dataset has been completely compiled for analysis. Once we have filled in the values that are missing, we will store the processed dataset using the Orange Save Data widget. Handling missing values is an essential part of data preparation, as it guarantees accuracy and prepares the data for analysis and modelling.



Info

1000 instances (no missing data)
7 features
No target variable.
No meta attributes.

Variables

☑ Show variable labels (if present)
□ Visualize numeric values
☑ Color by instance classes

Selection

☑ Select full rows

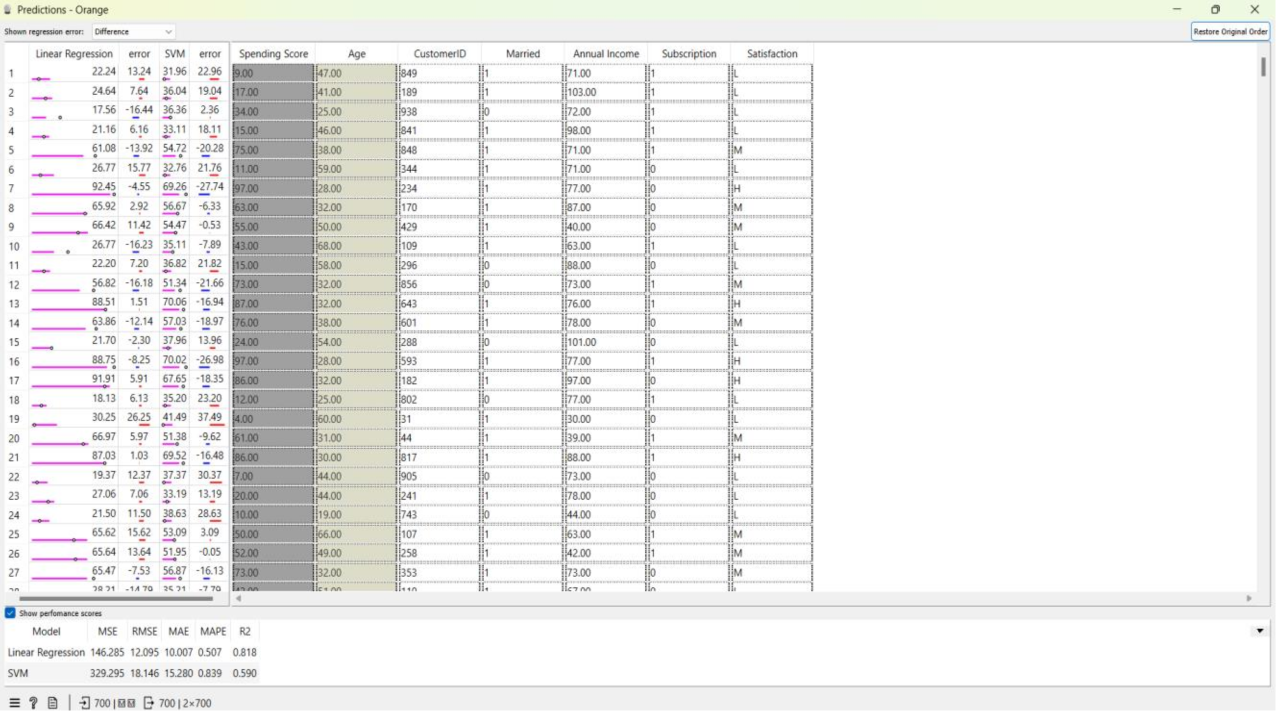
	CustomerID	Age	Married	Annual Income	Subscription	Satisfa
1	1	19	0	15	1	L
2	2	21	0	15	1	H
405	405	19	0	15	0	L
3	3	20	0	16	1	L
4	4	23	0	16	1	M
5	5	31	0	17	1	L
6	6	22	0	17	1	M
7	7	35	0	18	1	L
8	8	23	0	18	1	H
9	9	64	1	19	1	L
10	10	30	1	19	1	M
11	11	67	1	19	1	L
12	12	38.61	1	19	1	H
13	13	58	1	20	1	L
14	14	24	1	20	1	M
15	15	37	1	20	1	L
16	16	22	1	20	1	H
18	18	20	0	21	1	M
17	17	35	1	21	1	L
857	857	19	0	23	1	L
19	19	52	1	23	1	L

Linear Regression vs SVM

In the upcoming step, we will forecast the 'spending score' by employing the aforementioned method on the pristine Stella Pty. Ltd. dataset, which is devoid of any absent data.

To predict the spending score of customers for Stella Pty. Ltd. We have to compare. It’s 30% Dataset vs 20% Dataset (1vs2)

1. 30% Dataset

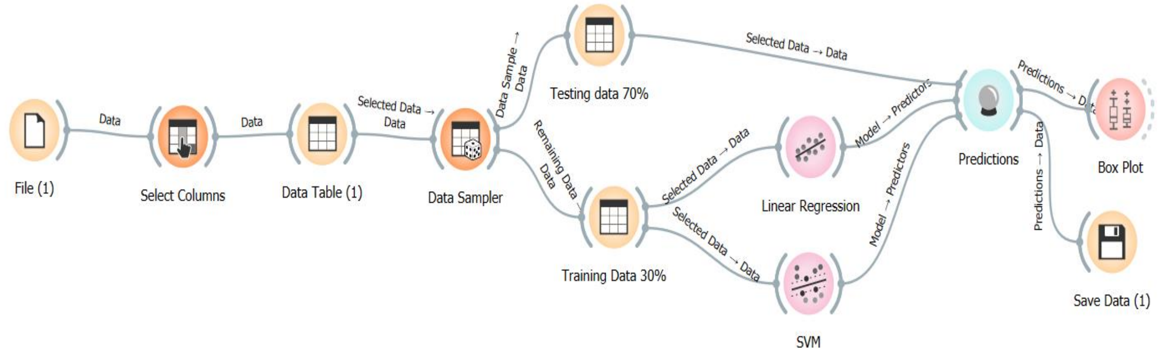


Shown regression error: Difference

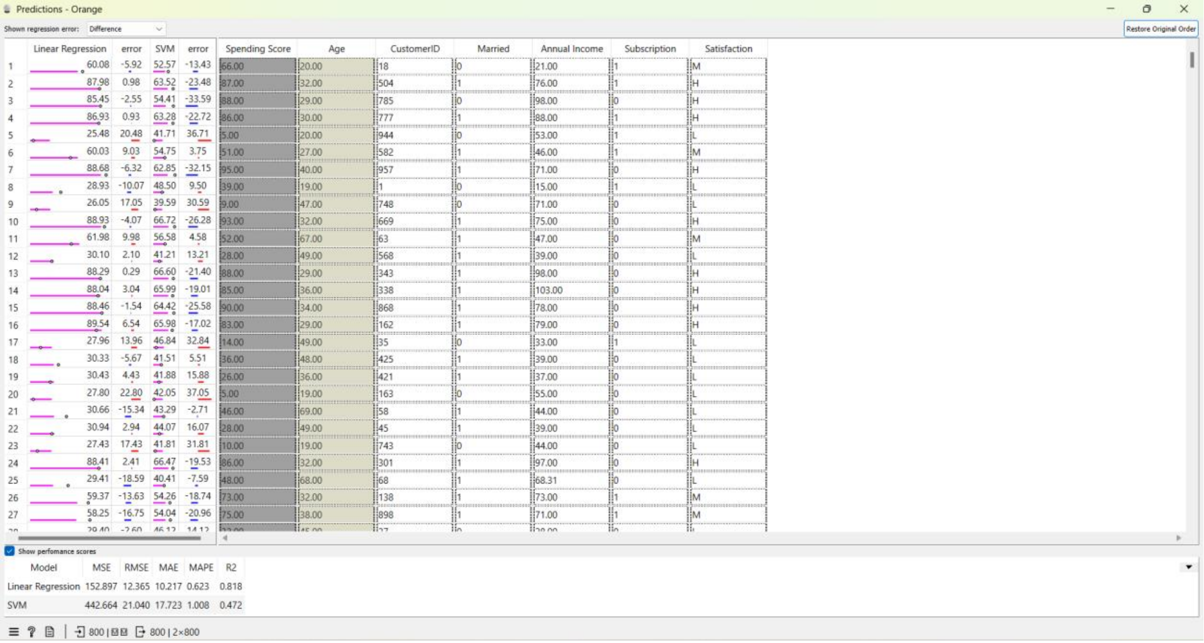
	Linear Regression	error	SVM	error	Spending Score	Age	CustomerID	Married	Annual Income	Subscription	Satisfaction
1	22.24	13.24	31.96	22.96	9.00	47.00	849	1	71.00	1	L
2	24.64	7.64	36.04	19.04	17.00	41.00	189	1	103.00	1	L
3	17.56	-16.44	36.36	2.36	34.00	25.00	938	0	72.00	1	L
4	21.16	6.16	33.11	18.11	15.00	46.00	841	1	98.00	1	L
5	61.08	-13.92	54.72	-20.28	75.00	38.00	848	1	71.00	1	M
6	26.77	15.77	32.76	21.76	11.00	59.00	344	1	71.00	0	L
7	92.45	-4.55	69.26	-27.74	97.00	28.00	234	1	77.00	0	H
8	65.92	2.92	56.67	-6.33	63.00	32.00	170	1	87.00	0	M
9	66.42	11.42	54.47	-0.53	55.00	50.00	429	1	40.00	0	M
10	26.77	-16.23	35.11	-7.89	43.00	68.00	109	1	63.00	1	L
11	22.20	7.20	36.82	21.82	15.00	58.00	296	0	88.00	0	L
12	56.82	-16.18	51.34	-21.66	73.00	32.00	856	0	73.00	1	M
13	88.51	1.51	70.06	-16.94	87.00	32.00	643	1	76.00	1	H
14	63.86	-12.14	57.03	-18.97	76.00	38.00	601	1	78.00	0	M
15	21.70	-2.30	37.96	13.96	24.00	54.00	288	0	101.00	0	L
16	88.75	-8.25	70.02	-26.98	97.00	28.00	593	1	77.00	1	H
17	91.91	5.91	67.65	-18.35	86.00	32.00	182	1	97.00	0	H
18	18.13	6.13	35.20	23.20	12.00	25.00	802	0	77.00	1	L
19	30.25	26.25	41.49	37.49	4.00	60.00	31	1	30.00	0	L
20	66.97	5.97	51.38	-9.62	61.00	31.00	44	1	39.00	1	M
21	87.03	1.03	69.52	-16.48	86.00	30.00	817	1	88.00	1	H
22	19.37	12.37	37.37	30.37	7.00	44.00	905	0	73.00	0	L
23	27.06	7.06	33.19	13.19	20.00	44.00	241	1	78.00	0	L
24	21.50	11.50	38.63	28.63	10.00	19.00	743	0	44.00	0	L
25	65.62	15.62	53.09	3.09	50.00	66.00	107	1	63.00	1	M
26	65.64	13.64	51.95	-0.05	52.00	49.00	258	1	42.00	1	M
27	65.47	-7.53	56.87	-16.13	73.00	32.00	353	1	73.00	0	M

☑ Show perfomance scores

Model	MSE	RMSE	MAE	MAPE	R2
Linear Regression	146.285	12.095	10.007	0.507	0.818
SVM	329.295	18.146	15.280	0.839	0.590



20% Dataset:

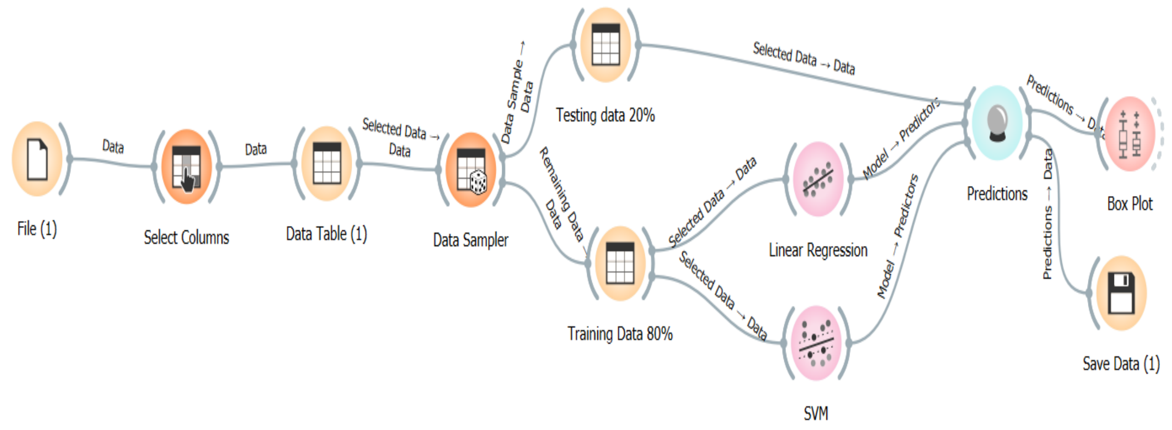


Shown regression error: Difference

	Linear Regression	error	SVM	error	Spending Score	Age	CustomerID	Married	Annual Income	Subscription	Satisfaction
1	60.08	-5.92	52.57	-13.43	66.00	20.00	18	0	21.00	1	M
2	87.98	0.98	63.52	-23.48	87.00	32.00	504	1	76.00	1	H
3	85.45	-2.55	54.41	-33.59	88.00	29.00	785	0	98.00	0	H
4	86.93	0.93	63.28	-22.72	86.00	30.00	777	1	88.00	1	H
5	25.48	20.48	41.71	36.71	5.00	20.00	944	0	53.00	1	L
6	60.03	9.03	54.75	3.75	51.00	27.00	582	1	46.00	1	M
7	88.68	-6.32	62.85	-32.15	95.00	40.00	957	1	71.00	0	H
8	28.93	-10.07	48.50	9.50	39.00	19.00	1	0	15.00	1	L
9	26.05	17.05	39.59	30.59	9.00	47.00	748	0	71.00	0	L
10	88.93	-4.07	66.72	-26.28	93.00	32.00	669	1	75.00	0	H
11	61.98	9.98	56.58	4.58	52.00	67.00	63	1	47.00	0	M
12	30.10	2.10	41.21	13.21	28.00	49.00	568	1	39.00	0	L
13	88.29	0.29	66.60	-21.40	88.00	29.00	343	1	98.00	0	H
14	88.04	3.04	65.99	-19.01	85.00	36.00	338	1	103.00	0	H
15	88.46	-1.54	64.42	-25.58	90.00	34.00	868	1	78.00	0	H
16	89.54	6.54	65.98	-17.02	83.00	29.00	162	1	79.00	0	H
17	27.96	13.96	46.84	32.84	14.00	49.00	35	0	33.00	1	L
18	30.33	-5.67	41.51	5.51	36.00	48.00	425	1	39.00	0	L
19	30.43	4.43	41.88	15.88	26.00	36.00	421	1	37.00	0	L
20	27.80	22.80	42.05	37.05	5.00	19.00	163	0	55.00	0	L
21	30.66	-15.34	43.29	-2.71	46.00	69.00	58	1	44.00	0	L
22	30.94	2.94	44.07	16.07	28.00	49.00	45	1	39.00	0	L
23	27.43	17.43	41.81	31.81	10.00	19.00	743	0	44.00	0	L
24	88.41	2.41	66.47	-19.53	86.00	32.00	301	1	97.00	0	H
25	29.41	-18.59	40.41	-7.59	48.00	68.00	68	1	68.31	0	L
26	59.37	-13.63	54.26	-18.74	73.00	32.00	138	1	73.00	1	M
27	58.25	-16.75	54.04	-20.96	75.00	38.00	898	1	71.00	1	M

☑ Show perfomance scores

Model	MSE	RMSE	MAE	MAPE	R2
Linear Regression	152.897	12.365	10.217	0.623	0.818
SVM	442.664	21.040	17.723	1.008	0.472



Observations:

Linear Regression:

- Linear Regression demonstrates constant performance across several data splits, including 70% and 80%. The R2 value remains unchanged at 0.818, showing a strong match for the data in both circumstances.
- Error Metrics: The error metrics (MSE, RMSE, MAE, MAPE) indicate that Linear Regression effectively captures the underlying trend in the data, as they are relatively low.

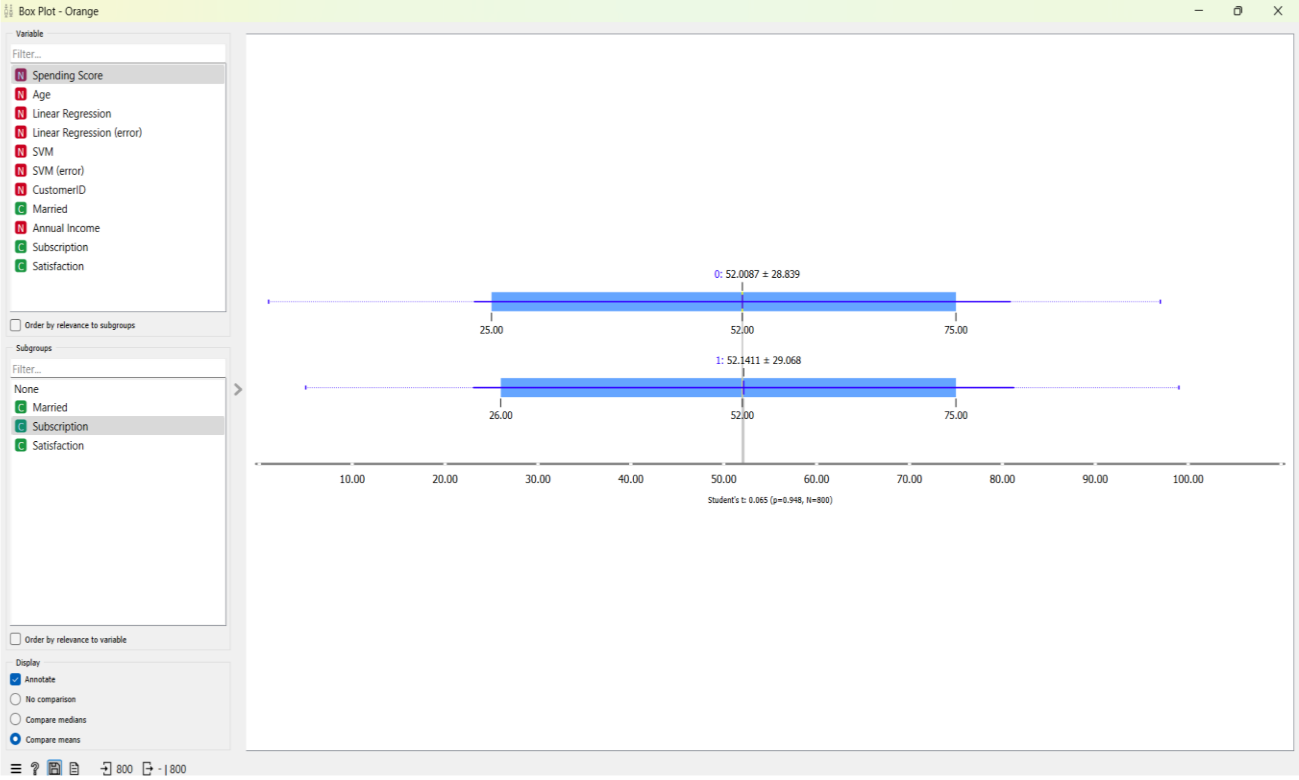
SVM:

- When comparing the error metrics of SVM and Linear Regression in both data splits, it is evident that SVM exhibits considerably greater error metrics. This implies that SVM is not as efficient in forecasting the expenditure score for this dataset.
- Lower R2 Scores: The R2 scores for Support Vector Machines (SVM) are 0.590 (70% data) and 0.472 (80% data), suggesting a worse fit in comparison to Linear Regression.

Based on the provided performance indicators, it can be determined that Linear Regression is a superior model for forecasting the spending score of customers for Stella Pty. Ltd. It consistently demonstrates reduced error metrics and greater R2 scores in comparison to

SVM. Hence, when it comes to forecasting spending scores, it is advisable to choose for the Linear Regression model rather than SVM for this particular dataset.

Boxplot 1:



Box Plot - Orange

Variable

Filter...

N Spending Score
N Age
N Linear Regression
N Linear Regression (error)
N SVM
N SVM (error)
N CustomerID
C Married
N Annual Income
C Subscription
C Satisfaction

☐ Order by relevance to subgroups

Subgroups

Filter...

None
C Married
C Subscription
C Satisfaction

☐ Order by relevance to variable

Display

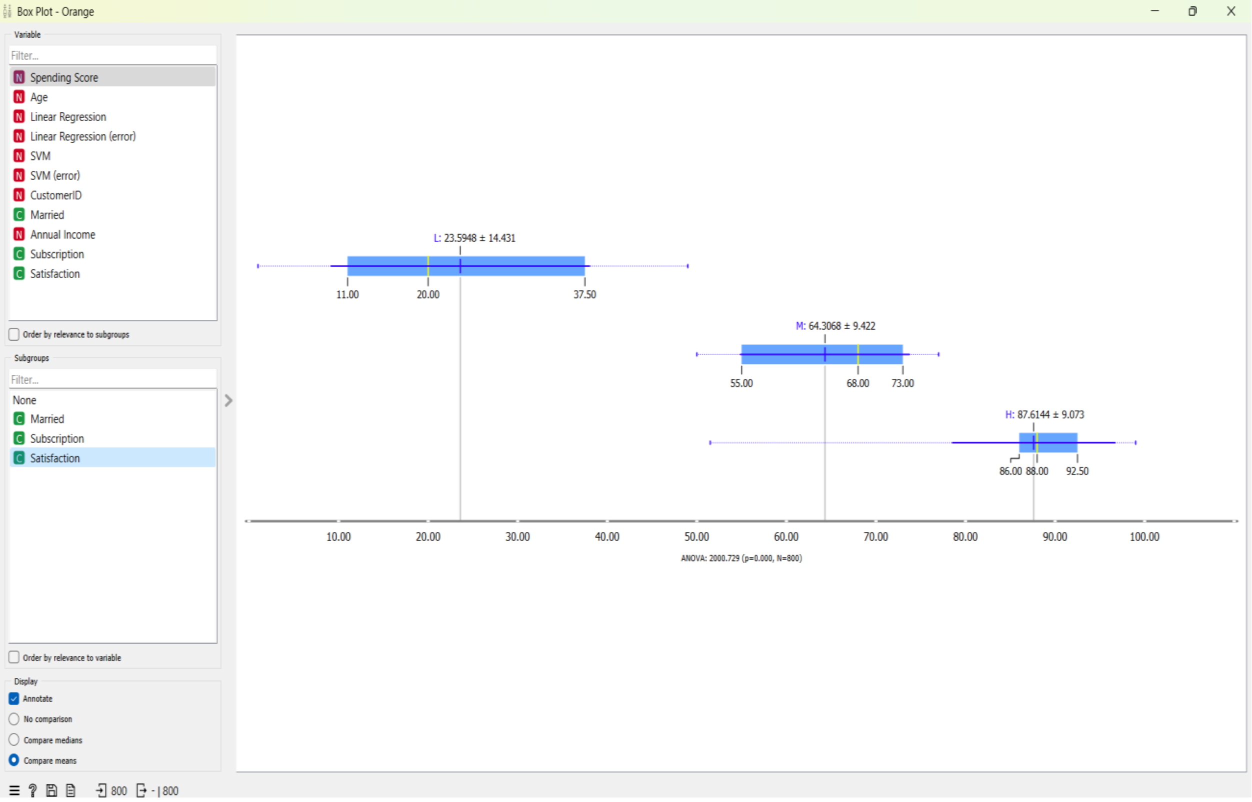
☑ Annotate
○ No comparison
○ Compare medians
● Compare means

800 - | 800

Observation:

The box plot illustrates the distribution of a variable known as "spending score" for two subgroups, denoted as "0" and "1". These divisions are most likely indicative of distinct subscription tiers or classifications. The median, which represents the central tendency, is the same for both groups and equals 52.00. The interquartile ranges (IQRs) exhibit a close resemblance, with values ranging from 25.00 to 75.00 for group "0" and 26.00 to 75.00 for group "1". This suggests that there is a similar level of variation in expenditure scores across the two subgroups. In addition, the averages (52.0087 for "0" and 52.1411 for "1") and standard deviations (28.839 for "0" and 29.068 for "1") are almost the same. A Student's t-test was performed, yielding a t-value of 0.065 and a p-value of 0.948, based on a sample size of 800. The elevated p-value indicates that there is no statistically significant distinction between the spending scores of the two subgroups.

Boxplot 2:



Box Plot - Orange

Variable

Filter...

- N Spending Score
- N Age
- N Linear Regression
- N Linear Regression (error)
- N SVM
- N SVM (error)
- N CustomerID
- C Married
- N Annual Income
- C Subscription
- C Satisfaction

☐ Order by relevance to subgroups

Subgroups

Filter...

- None
- C Married
- C Subscription
- C Satisfaction

☐ Order by relevance to variable

Display

☑ Annotate

○ No comparison

○ Compare medians

● Compare means

800 - | 800

Observation:

The above box plot illustrates the distribution of spending scores among three subgroups categorised by satisfaction levels: High (H), Low (L), and Medium (M). The primary observations are:

The grouping with high satisfaction has the highest median expenditure score of 92.5, suggesting that customers who are highly satisfied tend to have a higher spending tendency. This grouping also demonstrates the least amount of variation, characterised by a tight interquartile range and the absence of outliers. This indicates a steady pattern of high spending among consumers who are extremely satisfied. Conversely, the subgroup with low satisfaction exhibits the lowest median expenditure score of 37.5 and the greatest variability, characterised by a broad interquartile range and outliers indicating higher spending scores. The Medium satisfaction grouping is positioned in the middle, with a median spending score of 73.0 and a moderate level of variability. The ANOVA test yielded a statistically significant result (F=2000.729, p=0.000, N=800), indicating a notable disparity in expenditure scores among the three satisfaction classes.

Question 4:

Data Administration Approaches:

Stella Pty. Ltd. manages a regional chain of retail stores that offers a diverse range of women's apparel and accessories. As the business grows, the significance of data protection in DBMS becomes crucial. Below are two possible consequences of insufficient data protection and suggested security measures to minimise these risks:

Impact 1: Data Breach Leading to Financial and Reputational Damage

Discussion:

A data breach may occur if Stella Pty. Ltd.'s data is not sufficiently safeguarded. This breach potentially encompasses the illicit entry and pilferage of confidential client data, including names, addresses, purchase records, and payment particulars. Potential financial losses may result from legal penalties, compensation to impacted customers, and the expenses associated with breach remediation. Moreover, the company's standing would be compromised, leading to a decline in client confidence and probable revenue loss.

Recommended Security Feature: Encryption

Encryption is a highly effective security tool that can reduce the likelihood of data breaches. Through the process of encrypting sensitive data, Stella Pty. Ltd. guarantees that in the event an unauthorised entity gains entry to the database, the data will remain incomprehensible without the corresponding decryption key. The following are the fundamental components of encryption:

- Data-at-Rest Encryption: Utilises cryptographic algorithms to convert and secure data stored on discs and databases. The DBMS ensures that all stored data is encrypted, so safeguarding it against unauthorised access.
- Data-in-Transit Encryption ensures that data is encoded while being transferred across networks. Ensuring data security is of utmost importance throughout the transit of information between stores and the central database in order to prevent interception by malicious individuals.

Implementing encryption can greatly boost data security by rendering critical information unavailable to unauthorised users, thereby safeguarding against financial losses and reputational harm.

Impact 2: Data Integrity Issues Leading to Inaccurate Business Decisions

Discussion:

If Stella Pty. Ltd. does not have adequate data protection measures in place, the security and reliability of their data could be jeopardised. Data integrity concerns can occur due to inadvertent changes, deletions, or unauthorised adjustments to the database. Erroneous business reports and decisions can result from inaccurate or corrupted data, which can have a negative impact on inventory management, supplier relationships, and customer satisfaction. For instance, inaccurate inventory information might lead to either excessive stocking or insufficient stocking, both of which can have financial consequences.

Recommended Security Feature: Access Control

Implementing Access Control is crucial for preserving data integrity by controlling the individuals who have permission to access or alter the database. It guarantees that only individuals with proper authorisation can access particular data and capabilities within the database management system (DBMS). Essential elements of access control encompass:

- Role-Based Access Control (RBAC): It is a system where users are allocated specific roles according to their work responsibilities. Every position is assigned specific permissions that limit access to particular data and actions. Store managers may possess the ability to retrieve inventory data, but they do not have authorisation to view financial records.
- User Authentication: Prior to gaining access to the DBMS, users are required to confirm their identity, usually by providing passwords, undergoing biometric verification, or utilising multi-factor authentication (MFA). This method authenticates the user's identity and guarantees that only authorised users are granted access.
- Audit logs: They are essential for maintaining comprehensive records of database access and alterations. They play a crucial role in tracking unauthorised changes and promptly identifying the origin of data integrity problems. Systematic examinations of these records can promptly identify any dubious behaviours.

Enforcing access control mechanisms guarantees that only authorised persons can modify data, thereby maintaining its accuracy and reliability. Consequently, this facilitates well-informed and efficient company decision-making.

By implementing encryption and access control measures, Stella Pty. Ltd. may safeguard its data, maintain its reputation, and make informed business decisions based on reliable information.

Question 5:

Applications of Big Data and Analytics for Stella Pty. Ltd:

1. **Analysis of Customer data**

Overview:

Customer analytics is the process of examining customer data to obtain valuable information about their behaviour, interests, and purchasing habits. This encompasses demographic data, purchase records, browsing patterns, and customer input.

Advantages:

- Personalised Marketing: Stella Pty. Ltd. can customise marketing campaigns based on the specific interests and behaviours of customers, allowing for targeted marketing to individual customers or customer categories. Customised promotions and suggestions have the potential to enhance client involvement and devotion.
- Enhanced inventory management: Utilising information on client purchasing habits can enable accurate forecasting of demand for individual products. This enables the stores to maintain optimal inventory levels of popular items, hence minimising the chances of having excessive inventory or running out of stock.
- Enhanced Customer Experience: Analysing client feedback and behaviour enables the identification of areas for improvement in the purchasing experience, leading to an enhanced customer experience. This can result in enhanced client satisfaction and increased customer retention.

2. **Analysis of the supply chain**

Overview:

Supply chain analytics is the application of data to enhance the efficiency of acquiring, transporting, and storing commodities. The process involves evaluating the performance of suppliers, assessing delivery times, monitoring inventory levels, and analysing logistics costs.

Advantages:

- Stella Pty. Ltd. can optimise inventory levels across all outlets by analysing sales data and supplier delivery patterns. This guarantees that each business maintains an appropriate level of inventory to satisfy customer demand without incurring excessive expenses associated with maintaining inventory.
- Supplier Performance Management involves the use of data analytics to monitor and evaluate the performance of suppliers. This process helps to identify suppliers that can be relied upon and also highlights areas within the supply chain that require improvement. This can result in enhanced negotiation of conditions, decreased lead times, and enhanced quality of supplied goods.
- Data-driven analysis of transportation and logistics can provide valuable insights that enable the identification of opportunities to reduce costs. Opting for the most economical shipping methods and routes can effectively decrease total logistics costs.

Implementation Considerations:

In order for Stella Pty. Ltd. to optimally harness the potential of these big data and analytics applications, it is essential to undertake the subsequent steps:

1. Data Integration: Implement a resilient database management system that consolidates data from diverse sources, including sales transactions, customer profiles, supplier details, and shipping information. This solution is intended to supplant the present utilisation of Google Sheets and emails.

2. Data Quality Management: Guarantee the precision, coherence, and currency of the collected data. Enforcing data validation and cleansing procedures will ensure the maintenance of superior data quality.

3. Invest in analytical tools and platforms capable of processing massive datasets and generating valuable insights. Consider utilising tools such as Tableau, Power BI, or specialised software for customer analytics and supply chain analytics.

4. Training: Conduct training sessions to instruct staff on the utilisation of the new database system and analytical tools. This will enable them to make decisions based on data and optimise the advantages of the analytics tools.

Through the utilisation of customer and supply chain analytics, Stella Pty. Ltd. can acquire a more profound comprehension of its business operations, augment customer contentment, and increase overall efficiency and profitability.

Appendix:

Assumptions Based on ER Diagram and Relational Model:

Cardinality and Participation Constraints:

- A customer can buy multiple products, and a product can be bought by multiple customers, resulting in a many-to-many relationship represented by the Customer_Product table.
- A store can sell multiple products, and a product can be sold in multiple stores, resulting in a many-to-many relationship represented by the Store_Product table.
- A supplier can supply multiple products, and each product can be supplied by multiple suppliers, resulting in a many-to-many relationship in the relational model.

Attribute Representation:

- Composite attributes like Address (comprising Street, City, State, Postal_Code) are represented as individual columns in the relational model.

Business Rules:

- An employee works in only one store but may manage multiple stores. This is represented by the Works_in relationship and the Manager table.
- Each invoice is related to a customer, indicating that sales transactions are always associated with a customer.

References:

Fan, J., Han, F. and Liu, H., 2014. Challenges of big data analysis. National science review, 1(2),pp.293-314.

Gotseva, D., Gancheva, V. and Georgiev, I., 2011. Database backup strategies and recovery models.Challenges in higher education & research, pp.147-150.

Peker, M., Özkaraca, O. and Şaşar, A., 2018. Use of orange data mining toolbox for data analysis in clinical decision making: the diagnosis of diabetes disease. In Expert system techniques in biomedical science practice (pp. 143-167). Igi Global.

Mullins, C., 2002. Database administration: the complete guide to practices and procedures. Addison-Wesley Professional.

Bertino, E. and Sandhu, R., 2005. Database security-concepts, approaches, and challenges. IEEE Transactions on Dependable and secure computing, 2(1), pp.2-19.

Morabito, V., 2015. Big data and analytics. Strategic and organisational impacts.

Erevelles, S., Fukawa, N. and Swayne, L., 2016. Big Data consumer analytics and the transformation of marketing. Journal of business research, 69(2), pp.897-904.