MLM-2

PROJECT 1 REPORT

CUSTOMER SEGMENTATION

FOR INTERSTELLAR

TOURISM

Submitted To: **Prof. Amarnath Mitra**

Submitted By:

Adhyatik (311063)

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Prof. Amarnath Mitra for his exceptional guidance and mentorship throughout the learning process of machine learning concepts. His profound knowledge, clear explanations, and dedication to teaching have been instrumental in my understanding and application of these concepts.

Prof. Amarnath Mitra's teachings have been invaluable in shaping my ability to analyze and interpret data effectively. His expertise in unsupervised machine learning techniques, particularly clustering algorithms like K-Means, has been crucial in conducting this customer segmentation analysis for interstellar tourism.

I am grateful for the engaging and insightful lectures, where Prof. Mitra provided real-world examples and practical applications of machine learning concepts. His ability to break down complex topics into understandable components has been truly remarkable.

Furthermore, I would like to acknowledge Prof. Mitra's unwavering support and encouragement throughout this project. His invaluable feedback and suggestions have greatly contributed to the successful completion of this analysis and the generation of actionable insights.

I extend my heartfelt appreciation to Prof. Amarnath Mitra for imparting his knowledge and fostering a stimulating learning environment. His teachings have not only equipped me with the necessary skills but have also ignited a passion for exploring the vast potential of machine learning in solving complex business problems.

EXECUTIVE SUMMARY

This report presents a comprehensive customer segmentation analysis for the interstellar tourism industry using unsupervised machine learning techniques. The objective was to identify distinct customer segments based on their preferences, behaviors, and characteristics, enabling targeted marketing and operational strategies.

The analysis leveraged a dataset encompassing approximately 500,000 records, each representing an individual space travel experience. Unsupervised K-Means clustering, coupled with statistical techniques like the Kruskal-Wallis test and ANOVA, was employed to segment customers and determine the optimal number of clusters.

The findings revealed three distinct customer segments:

Cluster 0: Premium Segment - Willing to invest in high-priced experiences to destinations like Alpha Centauri, potentially during December, with a preference for medium travel distances and exhibiting medium customer satisfaction levels.

Cluster 1: Value-Driven Segment - Price-sensitive customers seeking affordable experiences to nearby destinations like Tau Ceti, possibly in January, with the lowest customer satisfaction levels.

Cluster 2: Adventurous Segment - Customers seeking immersive experiences at distant destinations like Proxima Centauri, likely in July, exhibiting the highest willingness to travel long distances, moderately priced preferences, and the highest customer satisfaction levels.

The analysis identified significant differentiating factors across these segments, including destination preferences, travel timing, distance to destination, price sensitivity, and customer satisfaction levels. These insights inform tailored marketing strategies, destination management, pricing models, transportation logistics, and continuous improvement efforts.

1.) <u>OBJECTIVES</u>

1.1) Customer Segmentation for Targeted Marketing Using Unsupervised Machine Learning Clustering Algorithms.

1.2.) Identify Number Of Optimal Segments.

1.3.) Determining the Characteristics of each Cluster.

2.) DESCRIPTION OF DATA

2.1. Data Source, Size & Shape

2.1.1. Data Source (Website Link):-

https://www.kaggle.com/datasets/anthonytherrien/interstellar-travel-customer-satisfaction-analysis

2.1.2. Data Size:- 74.6 MB

2.1.3. Data Description & Dimensions:- This dataset, titled "Interstellar Travel Customer Satisfaction Analysis," provides a comprehensive view of customer experiences in interstellar space travel. It encompasses approximately **500,000** records, each representing an individual space travel experience. The primary objective of this dataset is to understand and predict the Customer Satisfaction Score, a key indicator of service quality and customer experience in the burgeoning field of interstellar tourism and travel.

Number Of Observations Taken For Analysis:- **1,00,192**
Number Of Variables:- **18**

2.2. Description of Variables

2.2.1. Index Variable(s): Row ID

2.2.2. Categorical Variables or Features (CV):- 09

1. **Gender**: Gender of the traveler.
2. **Occupation**: Occupation of the traveler, such as Colonist, Tourist, Businessperson, etc.

3. **Travel Class**: Class of travel, e.g., Business, Economy, Luxury.
4. **Destination**: Interstellar destination.
5. **Purpose of Travel**: The primary purpose of travel, e.g., Tourism, Research, Colonization.
6. **Transportation Type**: Type of transportation, e.g., Warp Drive, Solar Sailing, Ion Thruster.
7. **Special Requests**: Any special requests made by the traveler.
8. **Loyalty Program Member**: Indicates if the traveler is a member of a loyalty program.
9. **Month**: Month of travel.

2.2.2.1. Categorical Variables or Features - Nominal Type:- 09

1. **Gender**
2. **Occupation**
3. **Travel Class**
4. **Destination**
5. **Purpose of Travel**
6. **Transportation Type**
7. **Special Requests**
8. **Loyalty Program Member**
9. **Month**

2.2.2.2. Categorical Variables or Features - Ordinal Type: 0

There are no Categorical Variables of Ordinal Type in this Dataset.

2.2.3. Non-Categorical Variables or Features: 8

1. **Age:** Age of the traveler.
2. **Distance to Destination (Light-Years):** The distance to the destination measured in light-years.
3. **Duration of Stay (Earth Days):** Duration of stay at the destination in Earth days.
4. **Number of Companions:** The number of companions accompanying the traveler.
5. **Price (Galactic Credits):** Price of the trip in Galactic Credits.

6. **Booking Date:** Date when the trip was booked.
7. **Departure Date:** Date of departure.
8. **Customer Satisfaction Score:** Indicator of service quality and customer experience.

2.3. Descriptive Statistics

2.3.1. Descriptive Statistics: Categorical Variables or Features

2.3.1.1. Count & Relative Frequency Statistics

1. **Gender**

Gender	Count	Relative Frequency
Male	58403	0.5829108112424146
Female	41789	0.41708918875758544

2. **Occupation**

Occupation	Count	Relative Frequency
Tourist	16768	0.16735867135100607
Businessperson	16756	0.1672389013094858
Scientist	16746	0.16713909294155221
Other	16666	0.1663406259980837
Colonist	16655	0.16623083679335676
Explorer	16601	0.1656918716065155

3. **Travel Class**

Travel Class	Count	Relative Frequency
Economy	54967	0.5486166560204407
Business	31828	0.3176700734589588
Luxury	13397	0.13371327052060045

4. Destination | **Count** | **Relative Frequency**

Destination	Count	Relative Frequency
Alpha Centauri	9265	0.09247245289045034
Zeta II Reticuli	9187	0.09169394762056851
Proxima Centauri	9167	0.09149433088470137
Gliese 581	9146	0.09128473331204089
Tau Ceti	9146	0.09128473331204089
Trappist-1	9127	0.0910950974129671
Barnard's Star	9101	0.09083559565633982
Kepler-22b	9019	0.09001716703928457
Epsilon Eridani	9019	0.09001716703928457
Lalande 21185	8967	0.08949816352603002
Exotic Destination 9	986	0.00984110507824976
Exotic Destination 5	919	0.009172389013094858
Exotic Destination 4	914	0.009122484829128074
Exotic Destination 3	910	0.009082561481954647
Exotic Destination 2	901	0.008992733950814437
Exotic Destination 8	898	0.008962791440434366
Exotic Destination 1	893	0.008912887256467583
Exotic Destination 7	880	0.008783136378153944
Exotic Destination 6	879	0.008773155541360588
Exotic Destination 10	868	0.008663366336633664

5. Purpose Of Travel

Purpose Of Travel	Count	Relative Frequency
Research	20352	0.2031299904183967
Other	20060	0.2002155860747365
Tourism	19935	0.1989679814755669
Colonization	19926	0.1988781539444267
Business	19919	0.1988082880868732

6. Transportation Type

Transportation Type	Count	Relative Frequency
Solar Sailing	25154	0.25105796870009583
Warp Drive	25085	0.2503692909613542
Ion Thruster	24977	0.24929136058767168
Other	24976	0.2492813797508783

7. Special Requests

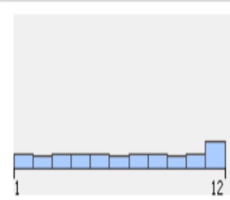
Special Requests	Count	Relative Frequency
Window Seat	20139	0.2010040721814117
Extra Space Suit	20101	0.20062480038326413
Special Meal	20017	0.19978641009262216
None	19991	0.1995269083359949
Other	19944	0.1990578090067071

8. Loyalty Status

Loyalty Status	Count	Relative Frequency
Yes	52117	0.520171271159374
No	48075	0.479828728840626

9. Month

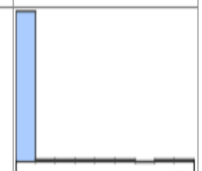
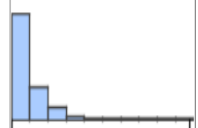
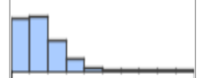
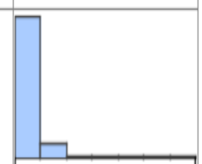
Month	Count	Relative Frequency
1	8744	0.08727243692111146
12	8596	0.08579527307569466
8	8548	0.08531619290961354
5	8544	0.08527626956244011
10	8525	0.08508663366336634
7	8496	0.084797189396359
3	8485	0.08468740019163207
4	8281	0.08265130948578729
9	8110	0.08094458639412329
11	8067	0.08051541041200894
6	8044	0.08028585116576174
2	7744	0.07729160012775471

Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No. -∞	Histogram
Month	1	6.5058	7	12	3.4582	-0.0041	-1.2112	8	0	0	

2.3.2. Descriptive Statistics: Non-Categorical Variables or Features

2.3.2.1. Measures of Central Tendency

2.3.2.2. Measures of Dispersion

Row ID	S Column	D Min	D Max	D Mean	D Std. de...	D Variance	D Skewness	D Kurtosis	D Overall ...	I No. mis...	I No. NaNs	I No. +∞s	I No. -∞s	D Median	I Row co...	Histogram
Age	Age	0	98	31.184	19.677	387.188	0.723	-0.127	3,124,363	0	0	0	0	28	100192	 0 98
Distance to D...	Distance to ...	0	1,795.31	8.302	23.635	558.637	19.638	841.539	831,550.93	26	0	0	0	2.7	100192	 0 1,795
Duration of St...	Duration of ...	0	391.5	34.067	33.317	1,110.015	1.85	4.557	3,412,536	22	0	0	0	22	100192	 0 392
Number of Co...	Number of C...	0	10	1.108	1.095	1.199	1.108	1.577	111,043	1	0	0	0	1	100192	 0 10
Price (Galactic...	Price (Galact...	-4,822.526	47,059.009	1,025.638	1,366.348	1,866,908.213	4.843	57.855	102,699,18...	60	0	0	0	630.696	100192	 -4.823 47.059
Customer Sati...	Customer S...	32.25	115	101.627	9.304	86.56	-2.372	7.715	10,178,498.88	37	0	0	0	102	100192	 32 115

2.3.2.3. <u>Correlation Statistics</u>

Correlation Matrix

Row ID	D Age	D Distance to Destina...	D Duration of Stay (...	D Number of Comp...	D Price (Galactic Cre...	D Customer Satisfac...
Age	1.0	0.0012018622331481...	-3.268359978797662...	-0.07617321187515...	-0.13291710909193...	-0.19332349866477...
Distance t...	0.00120186...	1.0	0.06848002091921744	0.001791694565926...	0.5703616094405309	-0.05419203578003...
Duration o...	-3.2683599...	0.06848002091921744	1.0	-0.06400337965526...	0.049832633893204...	-0.00128230287324...
Number of...	-0.0761732...	0.0017916945659265...	-0.064003379655267...	1.0	0.00871297329381 1...	-0.02143670845884...
Price (Gala...	-0.1329171...	0.5703616094405309	0.04983263389320477	0.00871297329381 1...	1.0	0.090895137117804...
Customer ...	-0.1933234...	-0.05419203578003211	-0.001282302873247...	-0.02143670845884...	0.090895137117804...	1.0

Row ID	S First col...	S Second column name	D Correlation value	D p value	I Degrees...
Row ID	Age	Distance to Destination (Light-Years)	0.0012018622331481806	0.7036319749300362	100190
Row1	Age	Duration of Stay (Earth Days)	-3.268359978797662E-4	0.9176038447543221	100190
Row2	Age	Number of Companions	-0.07617321187515685	8.22863333694245...	100190
Row3	Age	Price (Galactic Credits)	-0.13291710909193286	0.0	100190
Row4	Age	Customer Satisfaction Score	-0.19332349866477952	0.0	100190
Row5	Distance to ...	Duration of Stay (Earth Days)	0.06848002091921744	0.0	100190
Row6	Distance to ...	Number of Companions	0.0017916945659265738	0.5706322419283532	100190
Row7	Distance to ...	Price (Galactic Credits)	0.5703616094405309	0.0	100190
Row8	Distance to ...	Customer Satisfaction Score	-0.05419203578003211	4.78262973117410...	100190
Row9	Duration of ...	Number of Companions	-0.06400337965526706	1.94679859254818...	100190
Row10	Duration of ...	Price (Galactic Credits)	0.04983263389320477	0.0	100190
Row11	Duration of ...	Customer Satisfaction Score	-0.0012823028732476...	0.6848278575279976	100190
Row12	Number of C...	Price (Galactic Credits)	0.008712973293811595	0.00581658376333...	100190
Row13	Number of C...	Customer Satisfaction Score	-0.021436708458846328	1.15219608378159...	100190
Row14	Price (Galact...	Customer Satisfaction Score	0.09089513711780459	0.0	100190


corr = -1
corr = +1
corr = n/a
Age
D...
D...
N...
P...
C...
Age
"Age" (1/6)
Distance to Desti...
Duration of Stay ...
Number of Comp...
Price (Galactic Cr...
Customer Satisfa...

3.) ANALYSIS OF DATA

3.1. Data Pre-Processing

3.1.1. Missing Data Statistics and Treatment

3.1.1.1 Missing Data Statistics:

Field-wise Analysis:

1. **Distance To Destination :** 26 missing values
2. **Duration Of Stay :** 22 missing values
3. **Number Of Companions :** 1 missing values
4. **Price(Galactic Credits) :** 60 missing values
5. **Customer Satisfaction Score:** 37 missing values
6. **Month:-** 08 missing values

Distance To Destination, Duration Of Stay, Number Of Companions, Price, Customer Satisfaction Score All of these are NON-CATEGORICAL Variables and Month is a CATEGORICAL Variable.

Record-wise Analysis:

1. Total missing records: 154

3.1.1.2 <u>Missing Data Treatment</u>:

1. Since all the missing values are non-categorical, the treatment will focus mainly on non-categorical fields and also the treatment of the only categorical variable(Month).
2. Treatment of 'price' Variable: As 'price' is a non-categorical variable in numerical format, the Simple Imputer method is applied to all fields containing null values. The imputation strategy is 'mean', which replaces missing values with the mean value in each column. Similar Treatment is done for all the variables having missing values as all of them are Non-Categorical and have numerical values. Following the imputation process, no columns retain null values.
3. Treatment of ‘Month’ Variable: As 'Month' is a categorical variable in non-numerical format, the Simple Imputer method is applied to all fields containing null values. The imputation strategy is 'most frequent value', which replaces missing values with the mode value in each column.

3.1.1.3 <u>Missing Data Exclusion</u>:

1. Deletion of Records: If any Record(Rows) contains more than 50% of null values those Records are removed from the dataset.
2. Variable Deletion: If any Variable(Columns) contains more than 50% of null values persist, the corresponding variables are considered for removal.
3. Following the implementation of deletion procedures, the dataset achieves completeness.

3.1.2. Numerical Encoding of Categorical Variables or Features

Since all categorical variables in the dataset are nominal, we apply label encoding to transform them into numerical representations and the Encoding Schema is Alphanumeric Order.

3.1.3. Outlier Statistics and Treatment

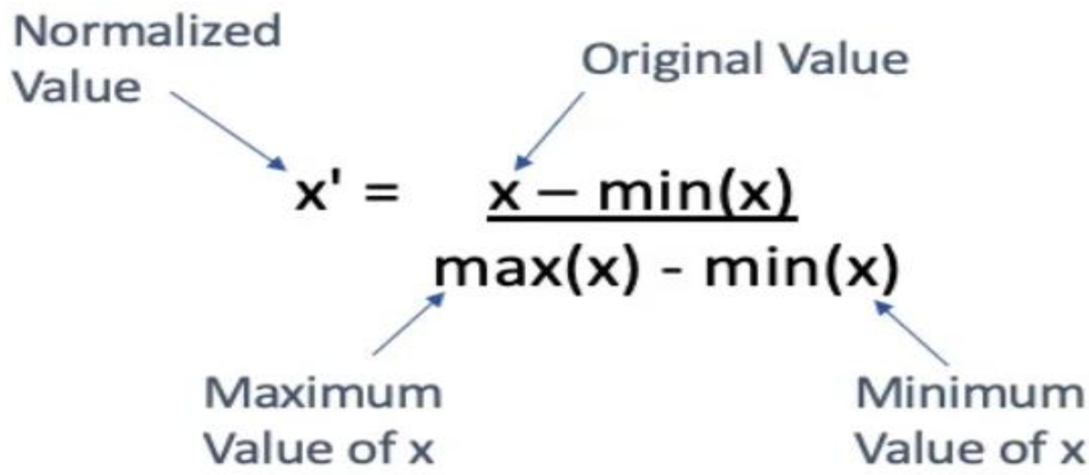
3.1.3.1.1. Outlier Statistics: Non-Categorical Variables or Features

Row ID	S Outlier c...	I Member ...	I Outlier c...	D Lower bo...	D Upper bo...
Row0	Age	100192	922	-24.5	83.5
Row1	Distance to ...	100192	10948	-8.69	17.07
Row2	Duration of ...	100192	5158	-43	101
Row3	Number of ...	100192	235	-3	5
Row4	Price (Galac...	100192	7295	-1,268.819	2,832.369
Row5	Customer S...	100192	6289	89.5	117.5

3.1.3.1.2. Outlier Treatment: Non-Categorical Variables or Features

1. Identification of Outliers: Box plots are utilized to visually inspect the presence of outliers within the dataset.
2. Outlier Detection: Outliers are identified in the following columns: ‘Distance To Destination’, ‘Duration Of Stay’, ‘Number Of Companions’, ‘Price’, ‘Customer Satisfaction Score’
3. Outlier Treatment: Min-Max Scaling normalization technique is applied to address outliers in the identified columns. Min-Max Normalization transforms x to x’ by converting each value of features to a range between **0 and 1**, and this is also known as **(0–1) Normalization.** If the data has negative values the range would have been between **-1 and 1**.

The formula for Min-Max Normalization is:



Formula for Min-Max Normalization

4. Min-Max Scaling normalizes the data and removes outliers. We utilize the interquartile range (IQR) from the 25th to the 75th percentile.

3.2. Data Analysis

3.2.1 Unsupervised Machine Learning Clustering Algorithm: K-Means

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It is an unsupervised learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

The working of the K-Means algorithm is explained in the below steps:

- **Step-1:** Select the number K to decide the number of clusters.
- **Step-2:** Select random K points or centroids. (It can be other from the input dataset).
- **Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters.
- **Step-4:** Calculate the variance and place a new centroid of each cluster.
- **Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
- **Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.
- **Step-7**: The model is ready.

3.2.2 Clustering Model Performance Evaluation: Silhouette Score

The silhouette coefficient or silhouette score k-means is a measure of how similar a data point is within-cluster (cohesion) compared to other clusters (separation).

- Select a range of values of k (say 1 to 10).
- Plot Silhouette coefficient for each value of K.

The equation for calculating the silhouette coefficient for a particular data point:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

- S(i) is the silhouette coefficient of the data point i.
- a(i) is the average distance between i and all the other data points in the cluster to which i belongs.
- b(i) is the average distance from i to all clusters to which i does not belong.

We will then calculate the average_silhouette for every k.

$$\text{AverageSilhouette} = \text{mean}\{S(i)\}$$

Then plot the graph between average_silhouette and K.

Points to Remember While Calculating Silhouette Coefficient:

- The value of the silhouette coefficient is between [-1, 1].
- A score of 1 denotes the best, meaning that the data point i is very compact within the cluster to which it belongs and far away from the other clusters.
- The worst value is -1. Values near 0 denote overlapping clusters.

3.2.3 Cluster Analysis: Base Model (K-Means)

3.2.3.1 Cluster Analysis with Categorical Variables or Features: Kruskal-Wallis Test

Categorical Variables: Kruskal-Wallis Test:

- Assumptions: The Kruskal-Wallis test is a non-parametric test and does not require the assumption of normality or homogeneity of variances. However, it assumes that the observations are independent and the dependent variable (categorical variable) is measured on at least an ordinal scale.

- Effect Size: In addition to the p-value, it is crucial to examine the effect size to gauge the practical significance of the differences between clusters, as statistical significance alone does not imply practical relevance.
- Post-hoc Tests: If the Kruskal-Wallis test indicates significant differences across clusters, post-hoc tests can be performed to identify which specific pairs of clusters differ significantly for the given categorical variable.
- Visualization: Visualizations such as bar plots, mosaic plots, or alluvial diagrams can aid in interpreting the distribution and patterns of categorical variables within and across clusters.

3.2.3.2 Cluster Analysis with Non-Categorical Variables or Features: Analysis of Variance (ANOVA)

Non-categorical Variables: One-way ANOVA:

- Assumptions: One-way ANOVA assumes normality of the dependent variable (non-categorical variable) within each cluster and homogeneity of variances across clusters. If these assumptions are violated, robust alternatives such as the Welch's ANOVA or non-parametric tests like the Kruskal-Wallis test can be considered.
- Effect Size: Similar to the Kruskal-Wallis test, effect size measures should be reported alongside the p-value to assess the practical significance of the differences between cluster means.
- Post-hoc Tests: When the ANOVA indicates significant differences, post-hoc tests (e.g., Tukey's HSD, Fisher's LSD for unequal variances) are required to identify which specific pairs of clusters differ significantly for the given non-categorical variable.
- Assumptions Validation: Before conducting ANOVA, it is essential to validate the assumptions of normality and homogeneity of variances (e.g., Levene's test,

Bartlett's test). If assumptions are violated, appropriate transformations or non-parametric alternatives may be required.

- Visualization: Box plots, violin plots, or interaction plots can help visualize the distribution and variability of non-categorical variables within and across clusters, aiding in the interpretation of the ANOVA results.

Multiple Testing Correction: When conducting multiple tests (e.g., testing multiple categorical and non-categorical variables), it is crucial to control the family-wise error rate (FWER) or the false discovery rate (FDR) to avoid an inflated Type I error (false positives). Techniques such as Bonferroni correction, Holm-Bonferroni method, or the Benjamini-Hochberg procedure can be employed to adjust the significance level accordingly.

Cluster Profiling: After identifying the distinguishing features and characteristics of each cluster, it is essential to profile the clusters by combining the insights from both categorical and non-categorical variables. This profiling can reveal patterns, preferences, or behaviors specific to each cluster, which can be leveraged for targeted marketing, product development, or operational strategies in the space travel and tourism industry.

Domain Knowledge Integration: Throughout the analysis process, it is crucial to collaborate with domain experts or stakeholders to ensure the interpretation of the cluster characteristics aligns with the business context and objectives. Their insights can validate the meaningfulness of the identified patterns, suggest alternative explanations, or highlight potential limitations or biases in the data or analysis methods.

Iterative Refinement: Cluster analysis is often an iterative process, where the initial results may prompt further exploration, variable selection, or adjustments to the clustering algorithm parameters. Continuous refinement and validation with domain experts can lead to more robust and actionable insights from the clustering analysis.

4.) OBSERVATIONS

4.1. Optimal Number of Clusters: Base Model (K-Means)

The Optimal Number Of Segments Or Clusters Comes Out To Be 3 Which Is a Non-Trivial Set Of Cluster When Unsupervised Machine Learning K-Means Clustering Algorithm Is Performed on the dataset of The Interstellar Space Travel And Tourism Because The Silhouette Score for K=3 is the Highest among all the Values for K(K=3,4,5) Performed And the Overall Mean For K=3 is 0.292 Which is the Highest Among all the Iterations and we are ignoring K=2 although having the highest Silhouette Score because we need to look for Non-Trivial clusters and K=2 is Trivial.

For K=2

Row ID	Mean Silhouette ...
cluster_0	0.345
cluster_1	0.384
Overall	0.359

For K=3

Row ID	D Mean Silhoue...
cluster_0	0.27
cluster_1	0.272
cluster_2	0.333
Overall	0.292

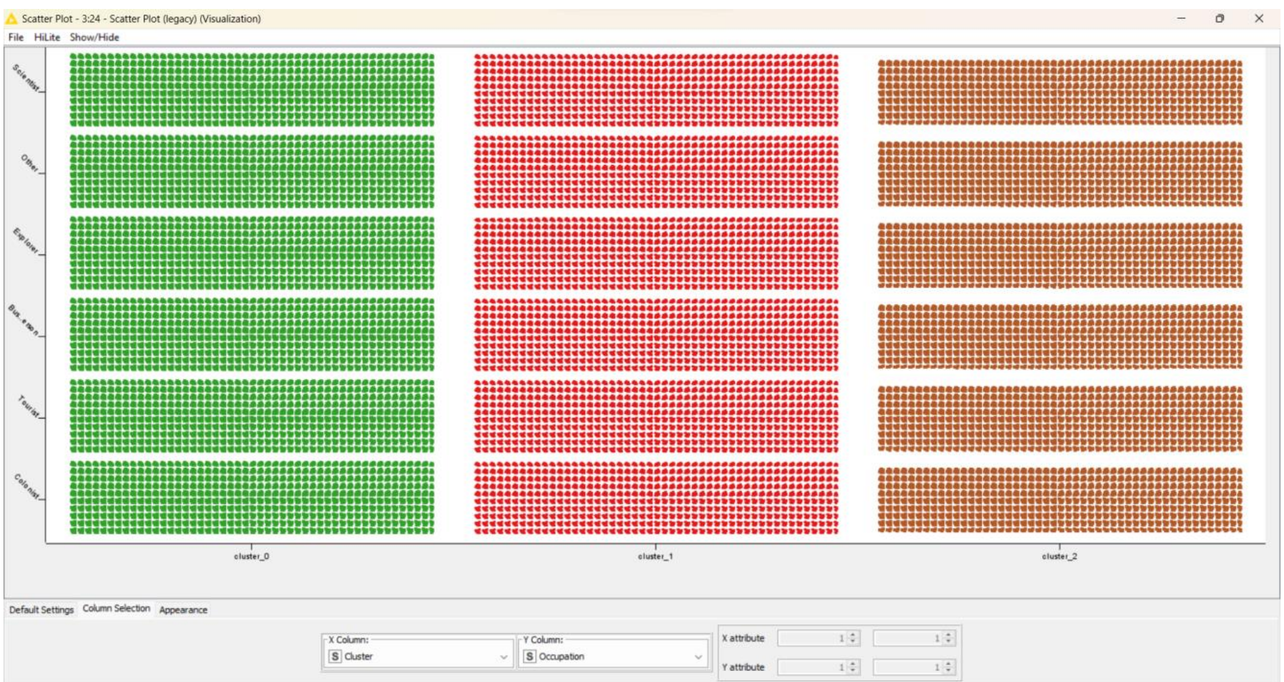
For K=4

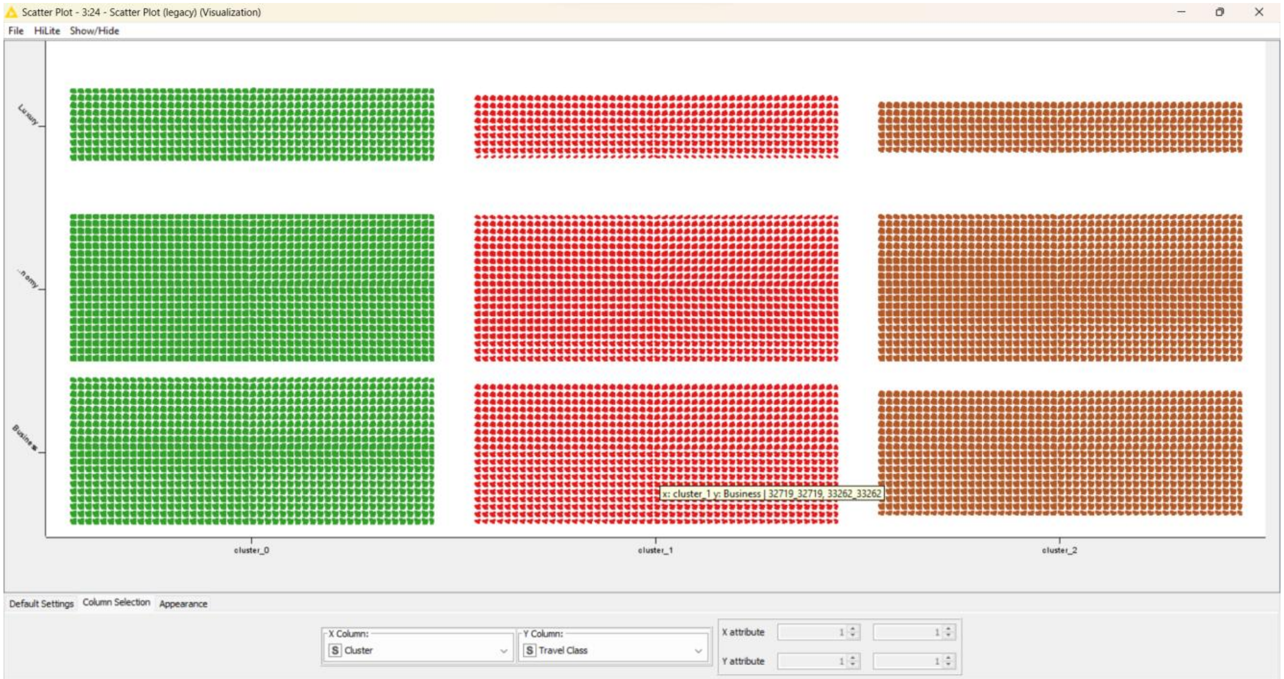
Row ID	D Mean Silhoue...
cluster_0	0.25
cluster_1	0.251
cluster_3	0.292
cluster_2	0.292
Overall	0.265

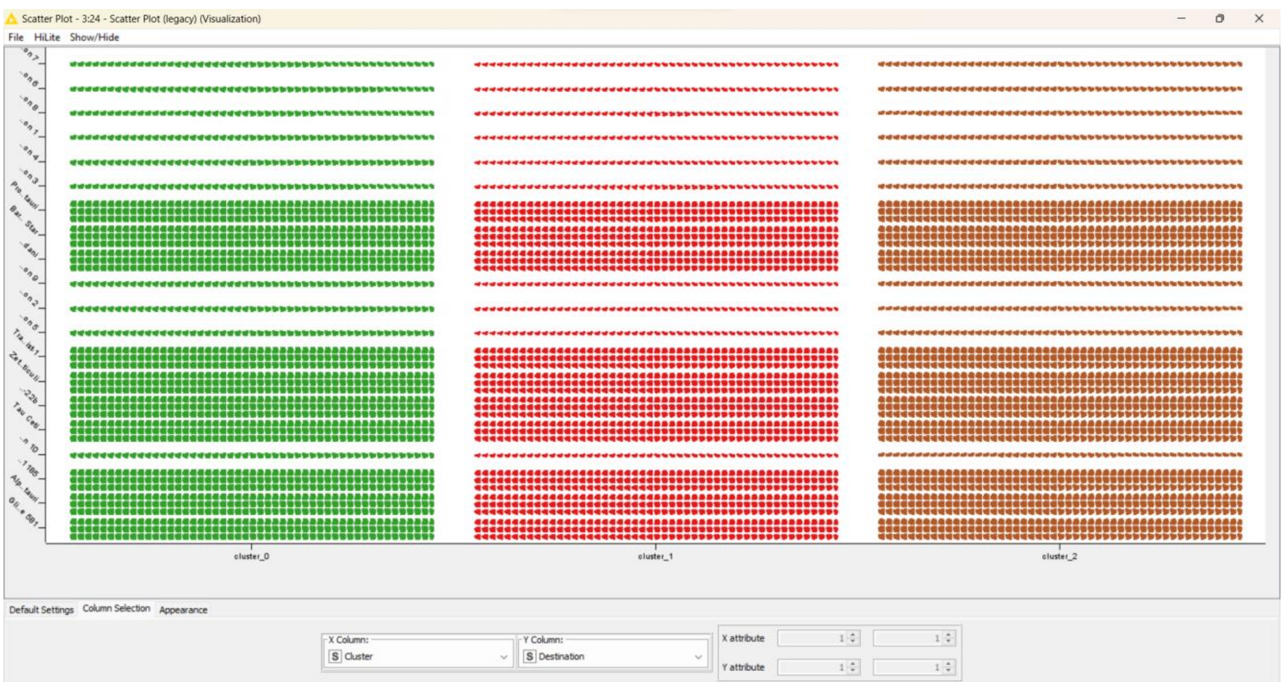
For K=5

Row ID	Mean Silhoue...
cluster_0	0.235
cluster_1	0.22
cluster_3	0.205
cluster_4	0.267
cluster_2	0.289
Overall	0.239

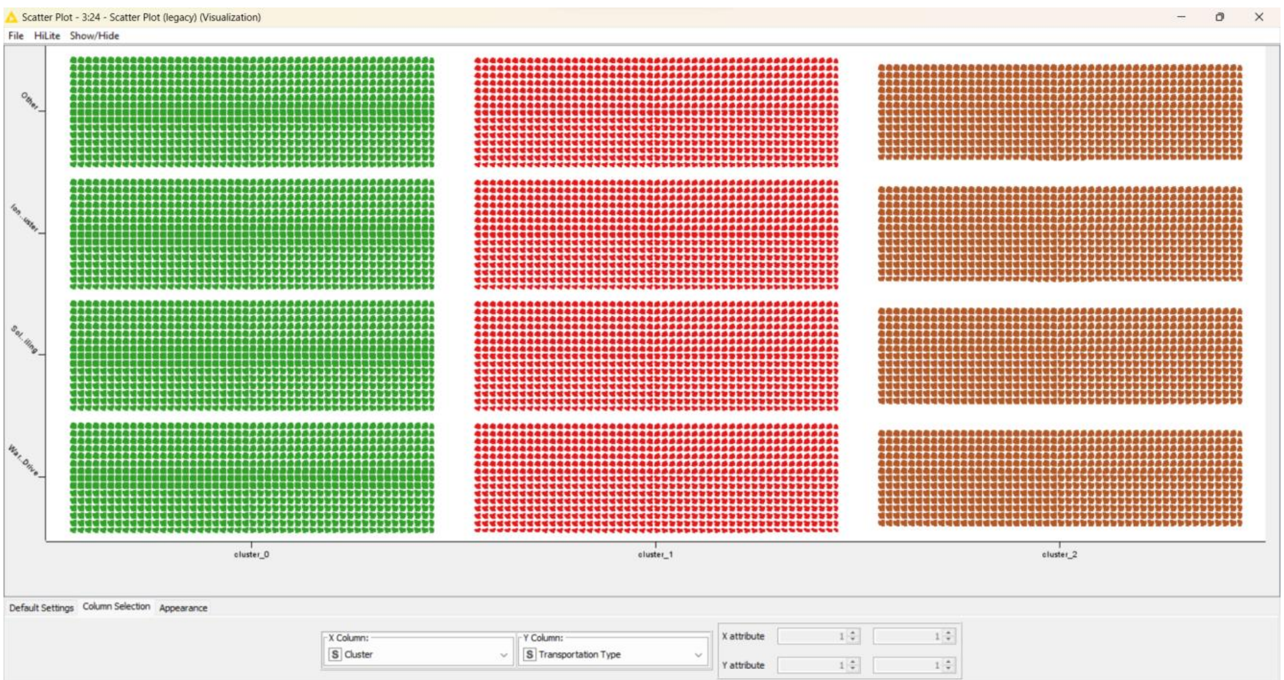
4.2 SCATTER PLOTS




Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
x: cluster_1 y: Business | 32719_32719, 33262_33262
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Travel Class
X attribute
Y attribute

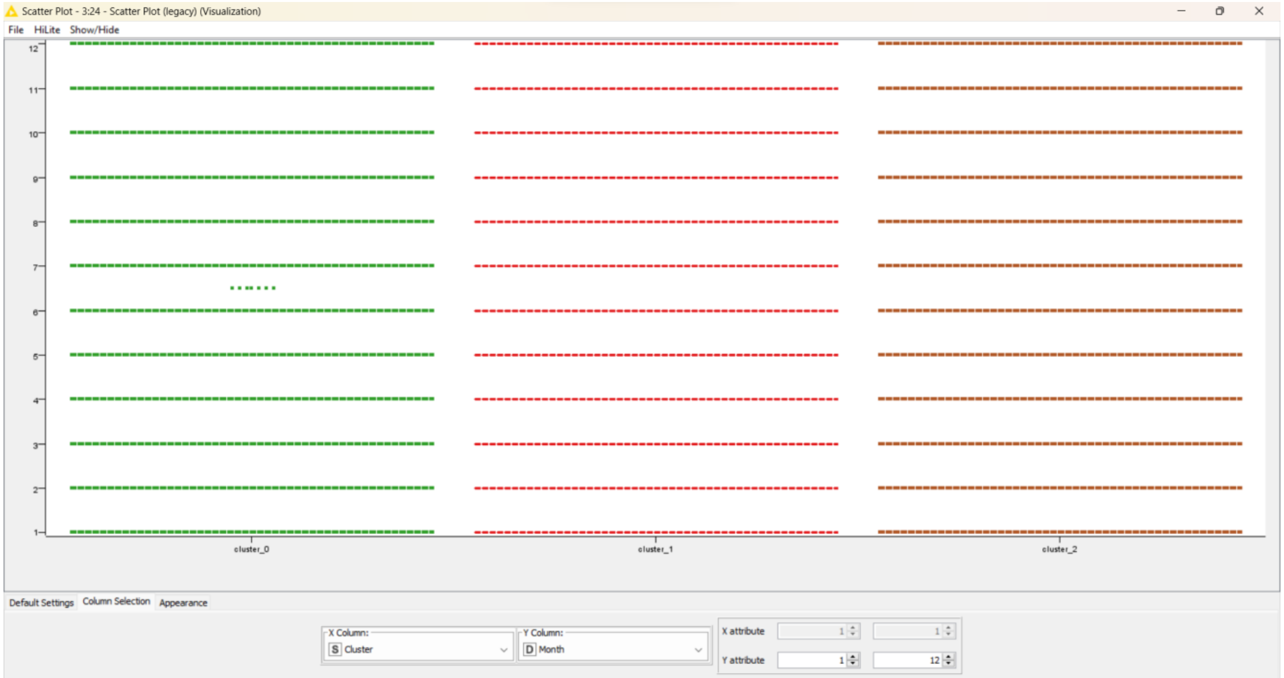

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Destination
X attribute
Y attribute

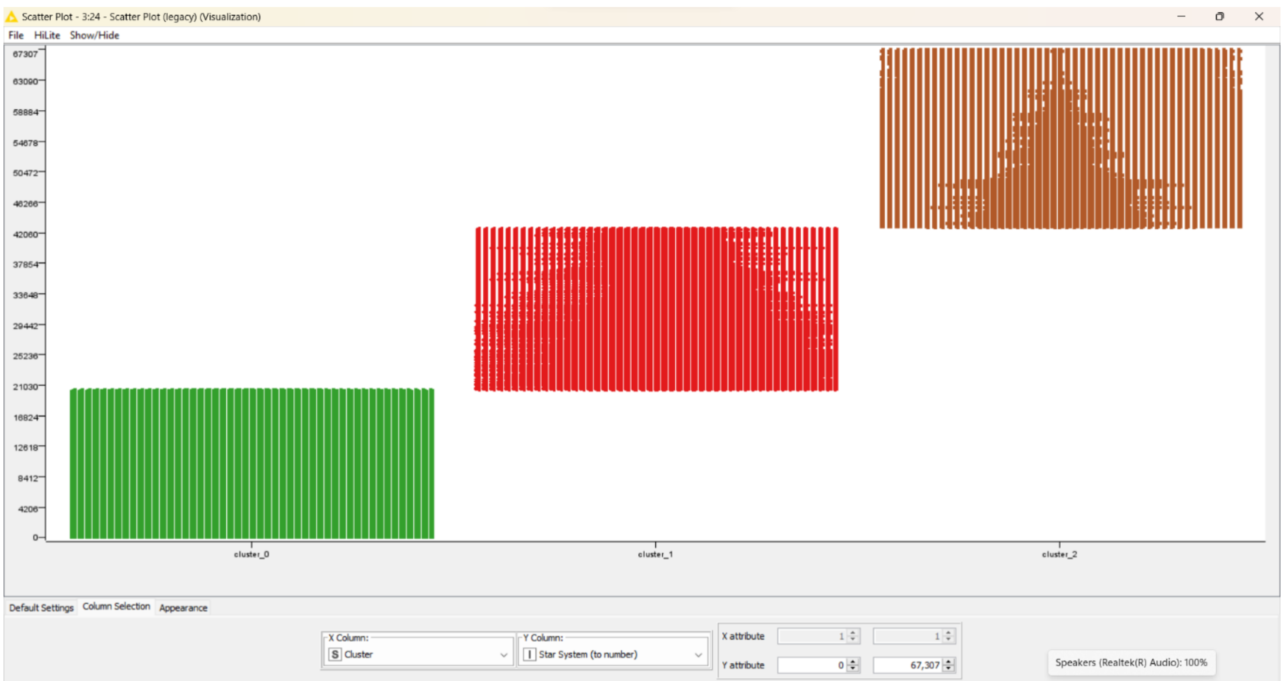

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Purpose of Travel
X attribute
Y attribute

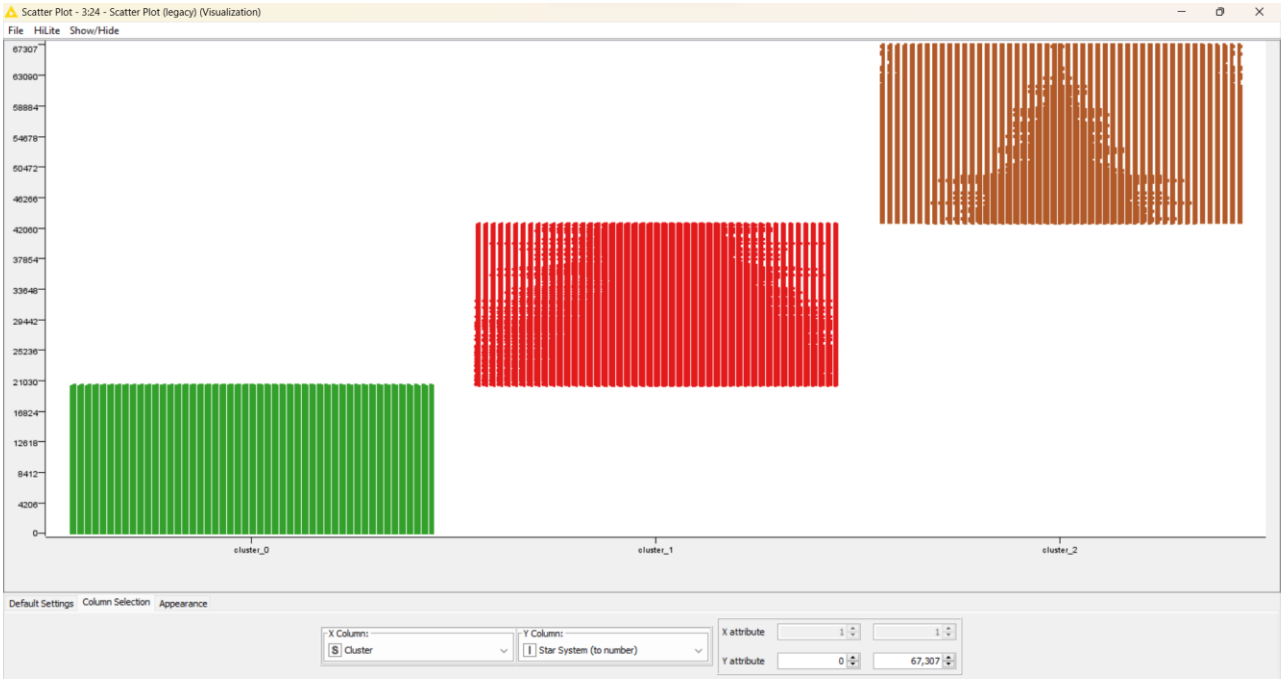

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Transportation Type
X attribute
Y attribute

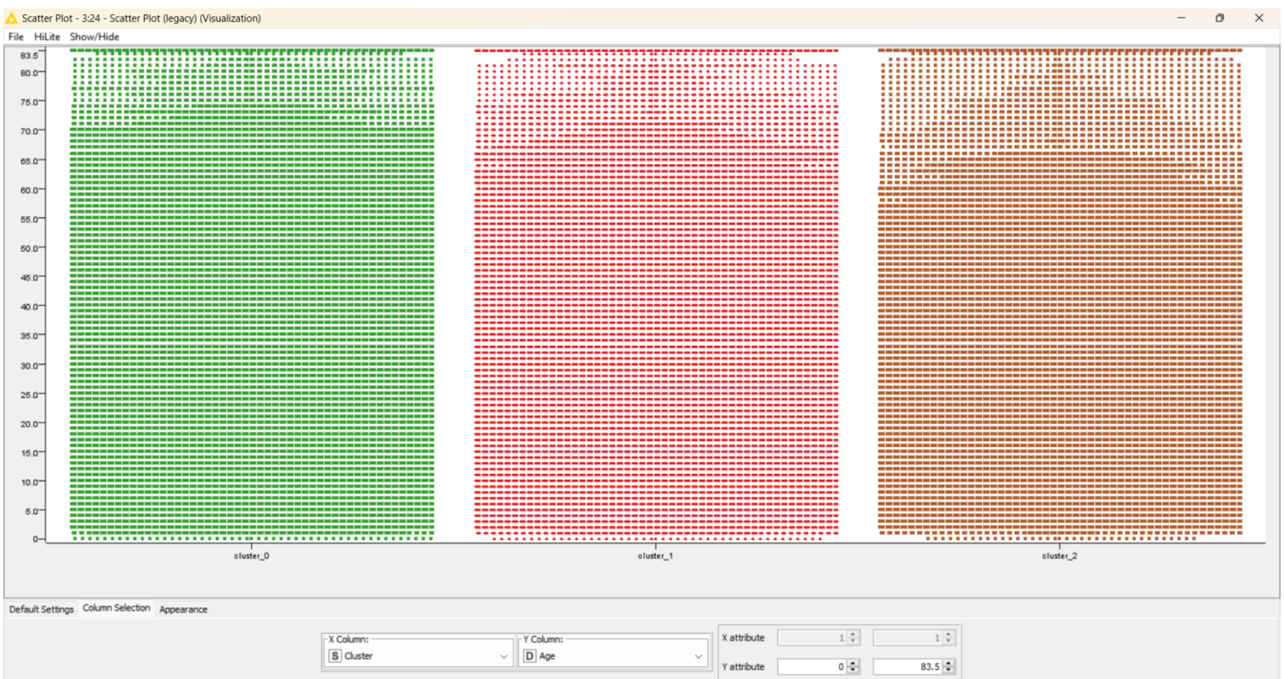

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
None
Other
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Special Requests
X attribute
Microsoft Store

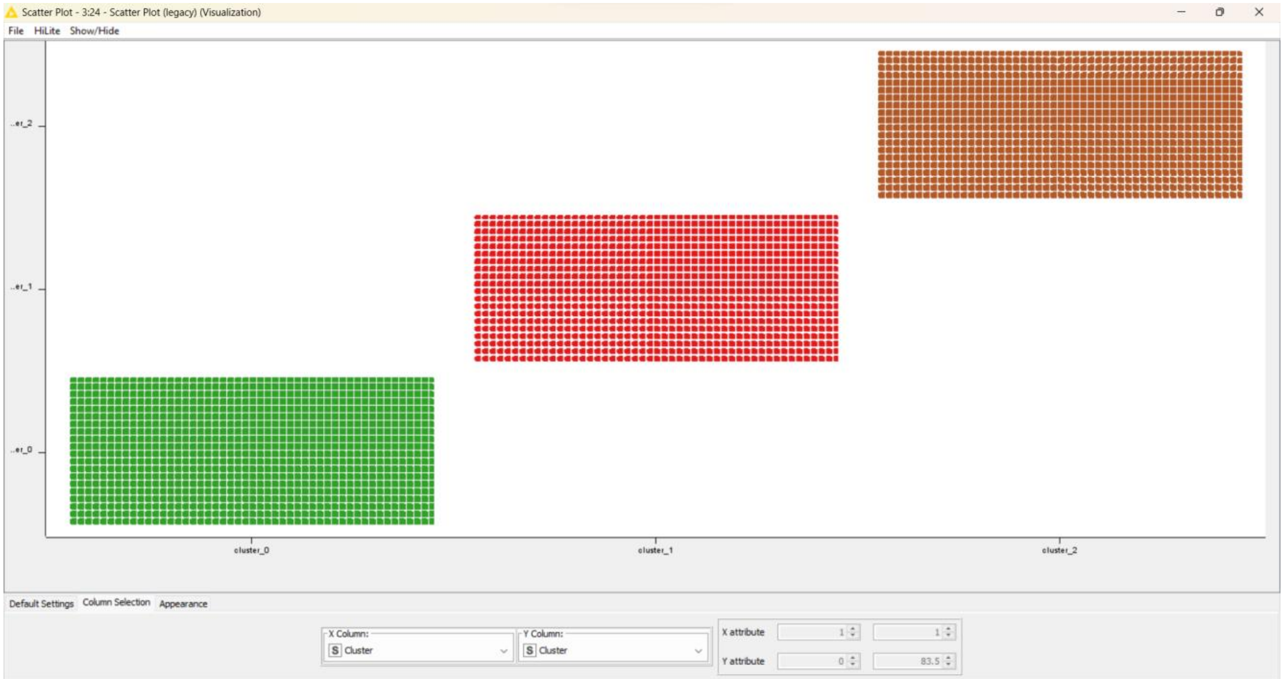

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
Yes
No
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Loyalty Program Member
X attribute
Y attribute

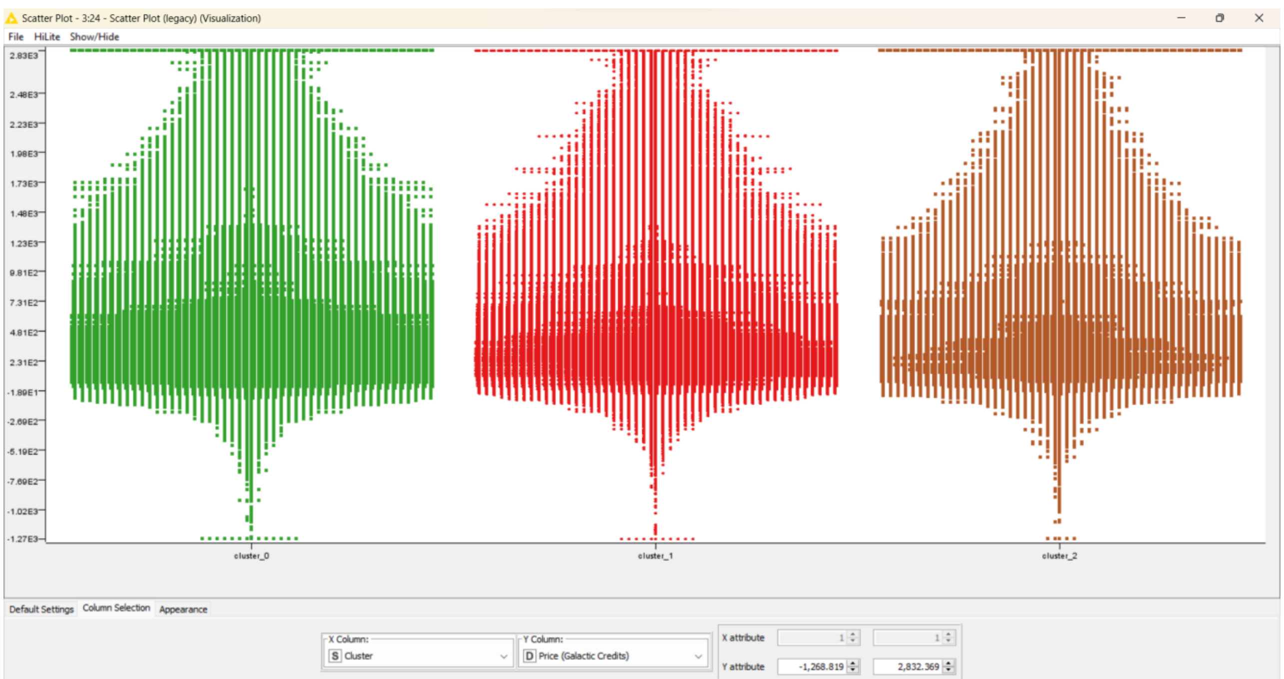

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column: Cluster
Y Column: Month
X attribute
Y attribute

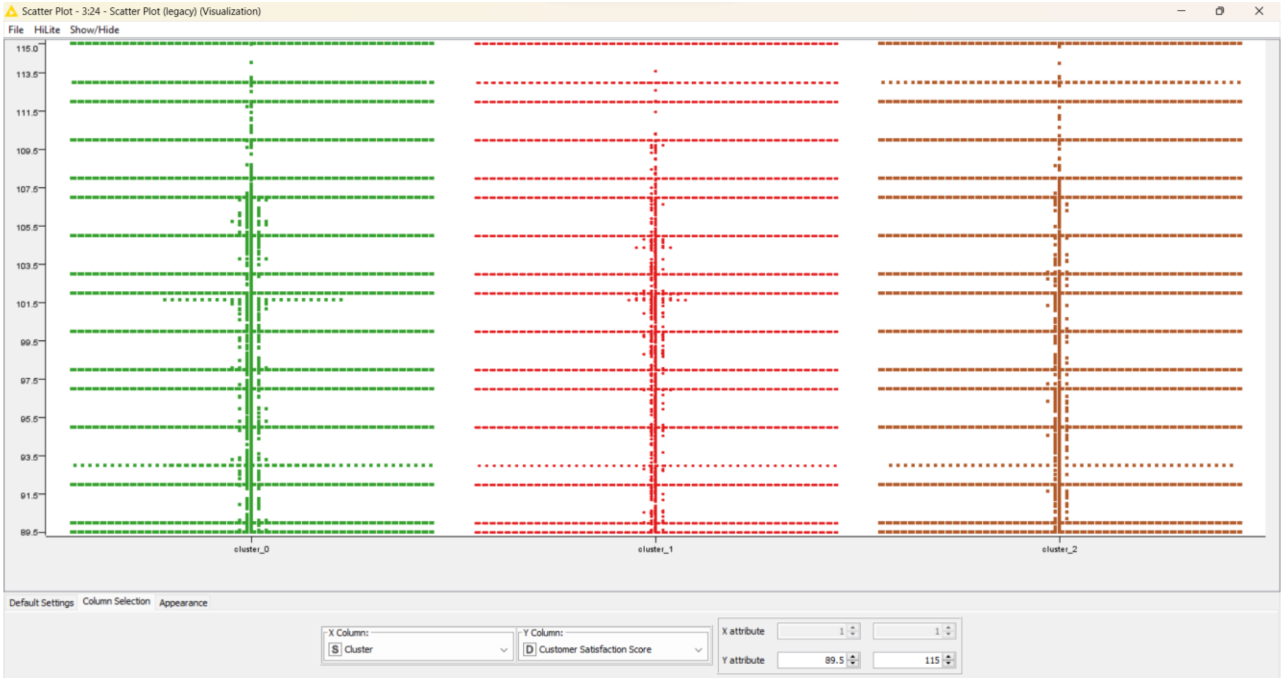

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column: Cluster
Y Column: Star System (to number)
X attribute
Y attribute
Speakers (Realtek(R) Audio): 100%

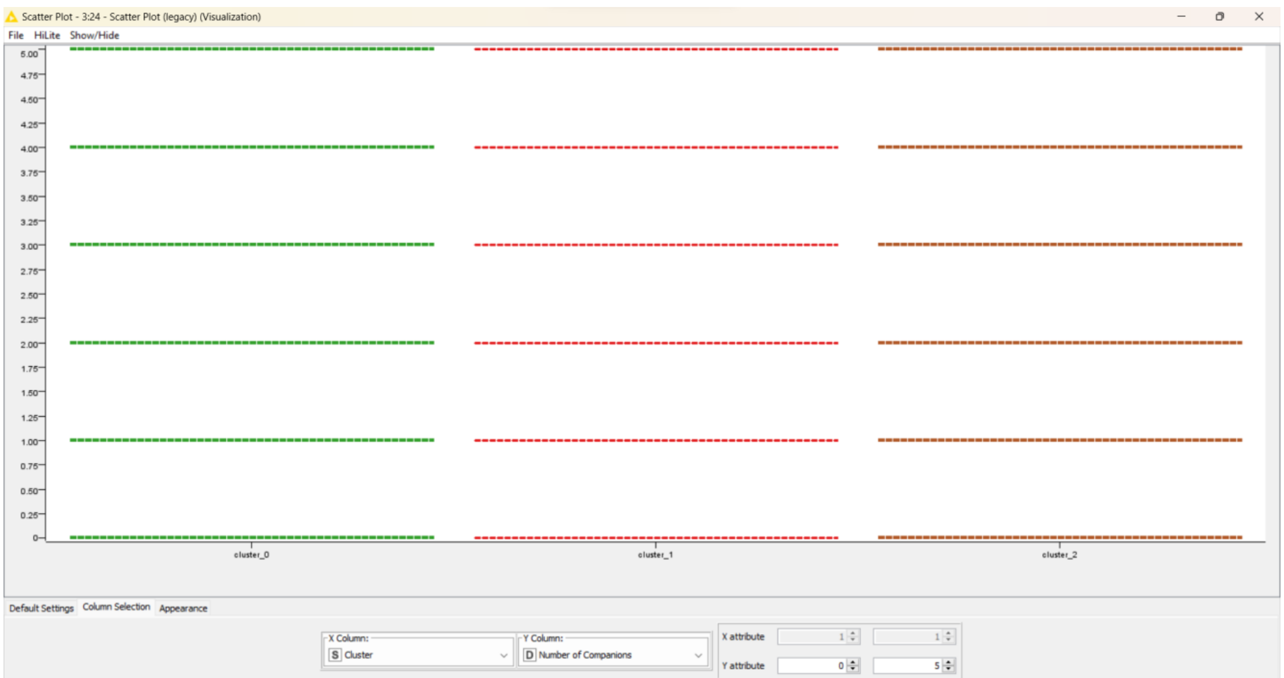

Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
67307
63090
58884
54678
50472
46266
42060
37854
33648
29442
25236
21030
16824
12618
8412
4206
0
cluster_0
cluster_1
cluster_2
Default Settings
Column Selection
Appearance
X Column:
S Cluster
Y Column:
I Star System (to number)
X attribute
Y attribute
67,307


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
83.5
80.0
75.0
70.0
65.0
60.0
55.0
50.0
45.0
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0
cluster_0
cluster_1
cluster_2
Default Settings
Column Selection
Appearance
X Column:
S Cluster
Y Column:
D Age
X attribute
Y attribute
83.5


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Cluster
X attribute
Y attribute


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Price (Galactic Credits)
X attribute
Y attribute
-1,268.819
2,832.369


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column: Cluster
Y Column: Customer Satisfaction Score
X attribute
Y attribute
89.5
115


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column: Cluster
Y Column: Number of Companions
X attribute
Y attribute
0
5


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Duration of Stay (Earth Days)
X attribute
Y attribute


Scatter Plot - 3:24 - Scatter Plot (legacy) (Visualization)
File HiLite Show/Hide
cluster_0
cluster_1
cluster_2
Default Settings Column Selection Appearance
X Column:
Cluster
Y Column:
Distance to Destination (Light-Years)
X attribute
Y attribute

4.3 Determining Characteristics From Cluster Analysis: Base Model (K-Means)

4.3.1 Categorical Variables or Features: Contributing or Significant | Non-Contributing or Non-Significant

The k-means clustering algorithm, coupled with the Kruskal-Wallis test, has proven to be an effective approach for identifying the significant categorical variables that contribute to the characterization of the three clusters in the space travel and tourism dataset. By analyzing the distributions of these variables across the clusters, valuable insights can be gained to inform business strategies and enhance customer experiences.

Contributing or Significant Categorical Variables:

In the context of the space travel and tourism dataset, the k-means clustering algorithm was employed to group the observations into distinct clusters. Following the cluster formation, the Kruskal-Wallis test, a non-parametric alternative to one-way ANOVA, was conducted to assess the significance of the categorical variables in differentiating the clusters. The analysis was performed at a 95% confidence level, with a significance threshold of 0.05 for the p-values.

The Kruskal-Wallis test results revealed that among the categorical variables, only two variables exhibited statistically significant differences in their distributions across the clusters. These contributing or significant variables were:

1. **Month**
2. **Destination**

Contributing or Significant Categorical Variables:

1. Month:

- The Month variable represents the temporal aspect of the space travel, potentially capturing seasonal patterns or preferences among customers.
- The significant differences in the distributions of Month across the clusters suggest that certain customer segments have distinct preferences or constraints regarding the timing of their space travel experiences.
- This information can guide space travel companies in optimizing their operational schedules, resource allocation, and promotional strategies based on the identified seasonal patterns or peak demand periods for each cluster.
- It can also inform the development of tailored offerings, such as seasonal packages or special events, catering to the preferences of specific customer segments.

2. Destination:

- The Destination variable represents the specific celestial destination or location visited by the customers during their space travel experience.
- The significant differences in the distributions of Destination across the clusters indicate that certain destinations hold varying degrees of appeal or popularity among different customer segments.
- This insight can assist space travel companies in identifying the most sought-after destinations for each cluster and optimizing their travel routes accordingly.
- It can also guide the development of targeted marketing campaigns, highlighting the unique features and attractions of popular destinations to resonate with the preferences of specific customer segments.
- Collaborations with destination management organizations and local tourism authorities can enhance the overall experience and offerings at popular destinations, catering to the interests of different clusters.

Non-Contributing or Non-Significant Categorical Variables:

While the Kruskal-Wallis test identified seven categorical variables as non-contributing or non-significant because as their p-value is greater than 0.1 in differentiating the three clusters, it is crucial to acknowledge that these variables should not be dismissed entirely. There are several considerations and potential avenues for further exploration:

1. **Interaction Effects**:

   - The non-contributing variables may exhibit significance when analyzed in combination with other variables or when considering interaction effects.
   - For example, the variable "Occupation" may not be significant on its own, but when combined with "Travel Class" or "Purpose of Travel," it could reveal interesting patterns or preferences specific to certain customer segments.
   - Exploring these interaction effects through advanced statistical techniques, such as two-way or higher-order ANOVA, or regression models with interaction terms, can uncover additional insights and nuances in the data.

2. **Alternative Clustering Techniques**:

   - The k-means algorithm is a widely used clustering technique, but it may not be the most suitable method for all datasets or variable types.
   - Alternative clustering algorithms, such as hierarchical clustering, DBSCAN, or model-based clustering, may reveal different patterns or highlight the significance of variables that were previously deemed non-contributing.
   - Experimenting with different clustering techniques and comparing the results can provide a more robust understanding of the underlying structure and relationships within the data.

3. **Domain Knowledge and Business Objectives**:

- While the statistical tests may deem certain variables as non-contributing, these variables should be evaluated in the context of domain knowledge and business objectives.
- For instance, the "Loyalty Program Member" variable may not contribute significantly to the cluster characterization, but it could be a critical factor in developing customer loyalty programs and retention strategies.
- Similarly, the "Gender" variable, although non-significant in this analysis, may be relevant for targeted marketing campaigns, product design, or addressing specific needs and preferences of different genders in the space travel industry.

4. **Data Quality and Completeness**:

- The non-contributing nature of certain variables may be influenced by data quality issues, such as missing values, inconsistencies, or measurement errors.
- Addressing these data quality concerns through data cleaning, imputation techniques, or collecting additional data can potentially reveal previously undetected patterns or relationships.

5. **Outlier Analysis and Robustness Checks**:

- Outliers or extreme values in the data can potentially influence the results of the statistical tests and clustering algorithms.
- Conducting robust outlier analysis and implementing appropriate techniques for handling outliers can improve the reliability and interpretability of the cluster characterization.

Month

Row ID	H-Value	p-value	Mean Rank of Group cluster_0	Median Rank of Group cluster_0	Mean Rank of Group cluster_1	Median Rank of Group cluster_1	Mean Rank of Group cluster_2	Median Rank of Group cluster_2
Row0	49,971.632	0.0	75,119.276	79,267	24,873.459	20,731	50,127.264	54,098.5

Destination

Row ID	H-Value	p-value	Mean Rank of Group cluster_0	Median Rank of Group cluster_0	Mean Rank of Group cluster_1	Median Rank of Group cluster_1	Mean Rank of Group cluster_2	Median Rank of Group cluster_2
Row0	68,276.289	0.0	32,954.384	32,819.5	32,824.278	32,819.5	82,976.674	81,101

4.3.2. <u>Non-Categorical Variables or Features: Contributing or Significant | Non-Contributing or Non-Significant</u>

In addition to the analysis of categorical variables, the characterization of the clusters in the space travel and tourism dataset was further enriched by examining the non-categorical variables using one-way ANOVA (Analysis of Variance). This statistical technique assesses whether the means of a continuous variable differ significantly across the identified clusters. The analysis was conducted at a 95% confidence level, with a significance threshold of 0.05 for the p-values.

<u>Contributing or Significant Non-Categorical Variables:</u>

The one-way ANOVA results revealed that three non-categorical variables exhibited statistically significant differences in their means across the clusters. These contributing or significant variables were:

1. Price (Galactic Credits)
2. Distance to Destination
3. Customer Satisfaction Score

1. Price (Galactic Credits):
- The Price variable represents the cost or financial investment associated with the space travel experience.
- The significant difference in the means of this variable across the clusters ($p\text{-value} < 0.05$) suggests that certain customer segments have distinct preferences or willingness to pay for their space travel experiences.
- This information can guide space travel companies in developing tailored pricing strategies, package offerings, and revenue management techniques to cater to the varying budget constraints and willingness to pay across different customer segments.
- It can also inform decisions regarding the allocation of resources, amenities, and service levels based on the price points preferred by each cluster.

2. Distance to Destination:

- The Distance to Destination variable represents the spatial distance, measured in light-years or other units, between the point of origin and the celestial destination visited by the customers.
- The significant difference in the means of this variable across the clusters (p-value < 0.05) indicates that different customer segments have varying preferences or constraints regarding the travel distance they are willing to undertake.
- This insight can guide space travel companies in optimizing their travel routes, allocating resources for longer or shorter journeys, and developing appropriate pricing strategies based on the distance factor.
- It can also inform the design and development of spacecraft or transportation modes optimized for different distance ranges, ensuring efficient and comfortable travel experiences for each customer segment.

3. Customer Satisfaction Score:

- The Customer Satisfaction Score variable represents the overall satisfaction level or rating provided by customers based on their space travel experience.
- The significant difference in the means of this variable across the clusters (p-value < 0.05) suggests that different customer segments have varying expectations, preferences, or perceptions of what constitutes a satisfactory space travel experience.
- This information can guide space travel companies in identifying the key drivers of customer satisfaction for each segment, enabling them to tailor their offerings, services, and amenities to meet or exceed the expectations of different customer groups.
- It can also inform continuous improvement efforts, service quality initiatives, and targeted customer experience strategies to enhance overall satisfaction and foster customer loyalty across all segments.

Non-Contributing or Non-Significant Non-Categorical Variables:

The remaining non-categorical variables, such as Age, Duration of Stay (Earth Days), Number of Companions, and Booking Date, were found to be non-contributing or non-significant in differentiating the clusters based on the one-way ANOVA results. Their p-values exceeded the 0.05 significance level, indicating that the means of these variables did not differ significantly across the clusters.

It is important to note that while the p-values for these variables were not statistically significant, some variables may exhibit potential differences in means across the clusters. However, these differences could be due to random chance or other factors, and further investigation or alternative statistical techniques may be required to assess their true significance.

Interpretation and Considerations:

1. Effect Size and Practical Significance:

- In addition to the p-values, it is crucial to examine effect size measures, such as eta-squared or omega-squared, to gauge the practical significance of the differences in means between the clusters.
- Effect sizes provide an understanding of the magnitude of the observed differences, ensuring that statistically significant results also translate to practical relevance in the context of the space travel and tourism industry.

2. Assumption Validation:

- One-way ANOVA relies on several assumptions, including normality of the variable within each cluster and homogeneity of variances across clusters.
- Violation of these assumptions can affect the validity of the results, and appropriate remedial measures, such as data transformations or robust alternatives (e.g., Welch's ANOVA or non-parametric tests), may be required.

3. Post-hoc Tests:

- When the one-way ANOVA indicates significant differences in means, post-hoc tests (e.g., Tukey's HSD, Bonferroni correction) are necessary to identify which specific pairs of clusters differ significantly for the given non-categorical variable.
- These pairwise comparisons provide valuable insights into the nature and direction of the differences, facilitating more targeted strategies and decision-making.

4. Interaction with Categorical Variables:

- The interpretation of the contributing and non-contributing non-categorical variables should be made in conjunction with the findings from the categorical variable analysis.
- Exploring potential interactions between categorical and non-categorical variables can uncover more nuanced patterns and relationships, leading to a more comprehensive understanding of the cluster characteristics.

5. Domain Knowledge and Business Context:

- As with the categorical variables, the interpretation of the non-categorical variable analysis should be guided by domain knowledge and the specific business objectives of the space travel and tourism industry.
- Collaboration with industry experts, stakeholders, and business analysts is crucial to ensure that the insights derived from the statistical analysis are actionable, practical, and aligned with the overall goals of enhancing customer experiences and optimizing operations.

ANOVA ANALYSIS TABLE

One-way analysis of variance (ANOVA)

Descriptive Statistics

Confidence Interval (CI) Probability: 95.0%

	Group	N	Missing	Missing Group	Mean	Std. Deviation	Std. Error	CI (Lower Bound)	CI (Upper Bound)	Minimum	Maximum
Age	cluster_0	32996	0	0	0.3732	0.2345	0.0013	0.3707	0.3757	0.0	1
Age	cluster_1	32776	0	0	0.3729	0.2337	0.0013	0.3704	0.3754	0.0	1
Age	cluster_2	34420	0	0	0.3729	0.2348	0.0013	0.3704	0.3754	0.0	1
Age	Total	100192	0	0	0.373	0.2343	0.0007	0.3715	0.3744	0.0	1
Distance to Destination (Light-Years)	cluster_0	32996	0	0	0.2999	0.3239	0.0018	0.2964	0.3034	0.0	1
Distance to Destination (Light-Years)	cluster_1	32776	0	0	0.2976	0.3219	0.0018	0.2941	0.3011	0.0006	1
Distance to Destination (Light-Years)	cluster_2	34420	0	0	0.3047	0.3258	0.0018	0.3012	0.3081	0.0006	1
Distance to Destination (Light-Years)	Total	100192	0	0	0.3008	0.3239	0.001	0.2988	0.3028	0.0	1
Duration of Stay (Earth Days)	cluster_0	32996	0	0	0.3222	0.2813	0.0015	0.3192	0.3252	0.0	1
Duration of Stay (Earth Days)	cluster_1	32776	0	0	0.3207	0.2819	0.0016	0.3177	0.3238	0.0	1
Duration of Stay (Earth Days)	cluster_2	34420	0	0	0.3228	0.2813	0.0015	0.3199	0.3258	0.0	1
Duration of Stay (Earth Days)	Total	100192	0	0	0.3219	0.2815	0.0009	0.3202	0.3237	0.0	1
Number of Companions	cluster_0	32996	0	0	0.2219	0.2158	0.0012	0.2195	0.2242	0.0	1
Number of Companions	cluster_1	32776	0	0	0.2206	0.2161	0.0012	0.2182	0.2229	0.0	1
Number of Companions	cluster_2	34420	0	0	0.2207	0.2171	0.0012	0.2184	0.223	0.0	1
Number of Companions	Total	100192	0	0	0.221	0.2163	0.0007	0.2197	0.2224	0.0	1
Price (Galactic Credits)	cluster_0	32996	0	0	0.5632	0.2103	0.0012	0.5609	0.5655	0.1496	1
Price (Galactic Credits)	cluster_1	32776	0	0	0.5038	0.2033	0.0011	0.5016	0.506	0.0	1
Price (Galactic Credits)	cluster_2	34420	0	0	0.5155	0.1983	0.0011	0.5134	0.5176	0.0	1
Price (Galactic Credits)	Total	100192	0	0	0.5274	0.2055	0.0006	0.5261	0.5287	0.0	1
Customer Satisfaction Score	cluster_0	32996	0	0	0.5021	0.2261	0.0012	0.4996	0.5045	0.0	1
Customer Satisfaction Score	cluster_1	32776	0	0	0.4934	0.2566	0.0014	0.4906	0.4961	0.0	1
Customer Satisfaction Score	cluster_2	34420	0	0	0.5487	0.2354	0.0013	0.5462	0.5512	0.0	1
Customer Satisfaction Score	Total	100192	0	0	0.5152	0.2409	0.0008	0.5137	0.5167	0.0	1

Levene Test

The Levene Test is used to test for the equality of variances.

	F	df 1	df 2	p-Value
Age	0.0304	2	100189	0.9701
Distance to Destination (Light-Years)	5.6971	2	100189	0.0034
Duration of Stay (Earth Days)	0.0275	2	100189	0.9729
Number of Companions	0.7352	2	100189	0.4794
Price (Galactic Credits)	166.7116	2	100189	0.0
Customer Satisfaction Score	217.6633	2	100189	0.0

ANOVA

	Source	Sum of Squares	df	Mean Square	F	p-value
Age	Between Groups	0.0025	2	0.0012	0.0226	0.9776
Age	Within Groups	5,501.5376	100189	0.0549		
Age	Total	5,501.5401	100191			
Distance to Destination (Light-Years)	Between Groups	0.8789	2	0.4395	4.1889	0.0152
Distance to Destination (Light-Years)	Within Groups	10,510.6161	100189	0.1049		
Distance to Destination (Light-Years)	Total	10,511.495	100191			
Duration of Stay (Earth Days)	Between Groups	0.081	2	0.0405	0.511	0.5999
Duration of Stay (Earth Days)	Within Groups	7,938.1119	100189	0.0792		
Duration of Stay (Earth Days)	Total	7,938.1928	100191			
Number of Companions	Between Groups	0.0341	2	0.017	0.3638	0.6951
Number of Companions	Within Groups	4,689.134	100189	0.0468		
Number of Companions	Total	4,689.1681	100191			
Price (Galactic Credits)	Between Groups	65.385	2	32.6925	786.2313	0.0
Price (Galactic Credits)	Within Groups	4,165.9848	100189	0.0416		
Price (Galactic Credits)	Total	4,231.3698	100191			
Customer Satisfaction Score	Between Groups	59.8869	2	29.9435	521.4895	0.0
Customer Satisfaction Score	Within Groups	5,752.7623	100189	0.0574		
Customer Satisfaction Score	Total	5,812.6492	100191			

5. MANAGERIAL INSIGHTS

Cluster Segmentation and Targeting:

Cluster 0: **Alpha Centauri (Destination), December (Month), Medium Distance to Destination, Highest Price, Medium Customer Satisfaction Score**

1. This segment represents customers willing to invest in premium space travel experiences to the Alpha Centauri destination, potentially during the month of December.
2. Marketing campaigns should emphasize the exclusivity and unique experiences offered at Alpha Centauri, aligning with the higher price point preferred by this segment.
3. Promotional efforts can highlight special events, luxury accommodations, or curated activities tailored to the preferences of this segment during the December travel period.

Cluster 1: Tau Ceti (Destination), January (Month), Lowest Distance to Destination, Lowest Price, Lowest Customer Satisfaction Score

1. This segment represents price-sensitive customers seeking affordable space travel experiences to nearby destinations like Tau Ceti, potentially in January.
2. Marketing strategies should focus on value-driven offerings, budget-friendly packages, and competitive pricing to attract this cost-conscious segment.
3. Emphasize the convenience and accessibility of the Tau Ceti destination, catering to the preference for shorter travel distances within this segment.
4. Implement targeted initiatives to address and improve customer satisfaction, as this segment exhibits the lowest satisfaction levels.

<u>Cluster 2:</u> Proxima Centauri (Destination), July (Month), Highest Distance to Destination, Medium Price, Highest Customer Satisfaction Score

1. This segment represents customers seeking immersive space travel experiences to distant destinations like Proxima Centauri, potentially during the month of July.
2. Marketing campaigns can highlight the adventure and exploration aspects of traveling to this distant destination, appealing to the sense of curiosity and adventure within this segment.
3. Offer moderately priced packages or flexible pricing options to cater to the medium price preference of this segment.
4. Leverage the high customer satisfaction levels within this segment by showcasing positive reviews, testimonials, and exceptional service standards.

<u>Destination Management and Experience Curation:</u>

1. Collaborate with destination management organizations and local tourism authorities to curate tailored experiences and activities at each destination, catering to the preferences of the respective customer segments.
2. Develop immersive educational programs, guided tours, or interactive exhibits at Alpha Centauri to align with the preferences of the premium segment (Cluster 0).
3. Enhance the on-site facilities and amenities at Tau Ceti to improve customer satisfaction levels for the price-sensitive segment (Cluster 1).
4. Invest in infrastructure and resources to support long-haul travel and extended stays at Proxima Centauri, catering to the preferences of the adventure-seeking segment (Cluster 2).

<u>Pricing and Revenue Management</u>:

1. Implement dynamic pricing strategies and yield management techniques to maximize revenue potential from the premium segment (Cluster 0) while maintaining exclusivity.
2. Offer bundled packages, discounts, and loyalty programs to attract and retain the price-sensitive segment (Cluster 1).
3. Explore flexible pricing options or modular offerings for the medium-priced segment (Cluster 2), allowing customers to customize their experiences based on preferences and budget.

Transportation and Logistics:

1. Optimize transportation routes and fleet allocation based on the varying distance preferences across the clusters, ensuring efficient travel to nearby destinations like Tau Ceti (Cluster 1) and longer-range travel capabilities for destinations like Proxima Centauri (Cluster 2).
2. Explore partnerships with transportation manufacturers and research institutions to develop advanced spacecraft or propulsion systems tailored to the different distance requirements of each segment.

Product Development and Innovation:

1. Invest in research and development to enhance existing spacecraft and transportation modes, catering to the diverse distance preferences across all clusters.
2. Explore the development of new propulsion systems, fuel-efficient engines, or sustainable transportation technologies to support long-haul travel to destinations like Proxima Centauri (Cluster 2) while minimizing environmental impact.
3. Collaborate with renowned designers, architects, and hospitality experts to develop innovative and immersive accommodation options at each destination, ranging from budget-friendly to luxury, catering to the preferences of all clusters.
4. Leverage emerging technologies, such as virtual reality and augmented reality, to create unique and engaging pre-travel experiences, enabling customers to explore their destinations virtually before their actual visit.

Strategic Partnerships and Collaborations:

1. Establish strategic partnerships with transportation manufacturers, aerospace companies, and research institutions to drive innovation in transportation modes and propulsion systems, aligning with the varying distance requirements of each cluster.
2. Collaborate with destination management organizations, local tourism authorities, and educational institutions to curate authentic and enriching experiences at each destination, tailored to the interests and preferences of different customer segments.
3. Partner with luxury hospitality brands and activity providers to create exclusive packages and personalized experiences for the premium segment (Cluster 0) visiting Alpha Centauri.
4. Foster collaborations with sustainable tourism organizations and eco-friendly initiatives to align with the preferences of environmentally conscious segments and promote responsible space travel practices.

Service and Operational Optimization:

1. Implement efficient check-in/check-out processes, streamlined services, and seamless transportation coordination for clusters with shorter travel distances (Cluster 1) to ensure a convenient and hassle-free experience.
2. Develop comprehensive training programs for staff and service personnel to deliver exceptional customer experiences tailored to the unique preferences and expectations of each cluster, addressing the low satisfaction levels within Cluster 1.
3. Leverage advanced data analytics and customer relationship management (CRM) systems to gain insights into customer preferences, enabling personalized recommendations and tailored services across all segments.
4. Optimize resource allocation, staffing, and on-site operations at each destination based on the varying duration preferences and peak travel periods identified within each cluster.

Continuous Improvement and Customer Experience:

1. Establish robust feedback mechanisms and customer satisfaction monitoring systems to continuously gather insights and preferences from customers across all clusters.
2. Implement targeted initiatives to address the low customer satisfaction levels within Cluster 1, such as service quality improvements, enhanced on-site experiences, or personalized attention.
3. Leverage the high customer satisfaction levels within Cluster 2 by encouraging positive word-of-mouth, testimonials, and referrals to attract new customers to the Proxima Centauri destination.
4. Foster a culture of continuous improvement and innovation within the organization, encouraging employees to contribute ideas and embrace change to maintain a competitive edge in the dynamic space travel and tourism industry.

By adopting a holistic approach that encompasses targeted marketing, destination management, pricing strategies, transportation logistics, and continuous improvement, space travel and tourism companies can effectively cater to the diverse preferences and characteristics of their customer segments. This comprehensive approach will enable the delivery of exceptional experiences, foster customer loyalty, and drive sustainable growth within the industry.

Group table - 4:37 - GroupBy

File Edit Hilite Navigation View

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	S Cluster	S Mode*(Destination)	S Unique concatenate with count*(Destination)	D Mode*(Month)	S Unique concatenate with count*(Month)	D Mean(Distance to Destination (Light-Years))	D Standard deviation(Distance to Destination (Light-Years))	D Mean(Price (Galactic Credits))	D Standard deviation(Price (Galactic Credits))	D Mean(Customer Satisfaction Score)	D Standard deviation(Customer Satisfaction Score)
Row0	cluster_0	Alpha Centauri	Gliese 581(4553), Lalan...	12	9(5230), 11(5415), 10(...	0.3	0.324	0.563	0.21	0.502	0.226
Row1	cluster_1	Tau Ceti	Alpha Centauri(4584), L...	1	3(5580), 5(5622), 4(54...	0.298	0.322	0.504	0.203	0.493	0.257
Row2	cluster_2	Proxima Centauri	Exotic Destination 5(16...	7	6.50582927413559(7), ...	0.305	0.326	0.516	0.198	0.549	0.235