**1. Word Count: 977 words**

**2. Introduction**

Provided with a dataset of 2,000 customers with information’s on their bookings and their interactions with the travel agency, our task is to identify distinct customer segments, from which we can suggest marketing strategies that match each target customer cluster. Explanatory Data Analysis will be conducted to uncover insights and important features of the variables. Customer Segmentation will be done using the K-means++ and Agglomerative Clustering techniques to identify potential clusters. The purpose of clustering is to group similar customers into the same group and efficiently target them through our marketing strategies. Recommendation of marketing strategies for each cluster will be given before reaching the Conclusion.

**3. Explanatory Data Analysis**

* The dataset includes 7 variables which are Gender, Marital Status, Age, Education, Income, Occupation and Settlement size.
* Regarding Gender, there are more Male customers at 60% compared to 40% that are female.

A blue and orange pie chart

Description automatically generated

* Regarding Age, most of the customers are around 25-55 years old, with most of customers are around 30-40 years old

A graph of a number of blue bars

Description automatically generated

* Regarding the level of Education of customers, High school graduates are highest at 43.8%, with University graduates comes second at 37.9%

A pie chart with text on it

Description automatically generated

* Regarding Income, its Mean is 137,516, larger than its Median at 133,004, the distribution of Income is right- skewed. Most of the income is around 90,000 to 170,000$.

A graph of income histogram

Description automatically generated

* Regarding Occupation, Unemployed/ Unskilled group accounts for the highest with 49.6%, and the second highest group is Skilled employee / Official at 39.6%.

A pie chart with text on it

Description automatically generated

* Regarding Settlement Size, most of them are Small city at 56.5% while Big city accounts for 39.9%.

**A pie chart with text

Description automatically generated**

* Regarding correlation between variables, overall, the correlation between the variables is high as no correlation is under 0.35. Marital Status-Occupation are highly correlated at 0.90 while Marital Status-Education, Marital Status-Settlement Size, Education-Occupation are high at 0.83, 0.79, 0.75 respectively.

**A screenshot of a graph

Description automatically generated**

**4. Customer Segmentation**

* The two numeric variables, that are Age and Income are standardized using StandardScaler.  Clustering models are distance-based algorithms as it forms a cluster using a distance metric. Variables with higher variation will have a higher impact on Clustering. Standardization helps to make the weight of each variable equal by converting to a unitless measure or relative distance. It allows all variables to be considered by the algorithm with equal importance.
* Using the Elbow method, the number of clusters can either be 2-5 based on the figure below

A graph with a line

Description automatically generated

* Combined with the test of Silhouette plots for 3 different value of 2; 3; 4, we can confirm that the number of clusters should be 2 as Silhouette’s average number is 0.54, closest to 1 when number of clusters is 2.

A graph of a graph

Description automatically generated with medium confidence

* K-means ++ and Agglomerative Clustering methods are used on the scaled dataset to cluster customers based on their information. With the number of clusters being 2, customers belong either to Cluster 0 or Cluster 1.

**A black text on a white background

Description automatically generated**

**A table of clustering information

Description automatically generated**

* Cluster 0 for K-means ++ is a cluster of late middle aged (age 48) Female, who is not single. She graduates from University and is a skilled employee. She lives in a big city and earns a high-income, average at $173,461.
* Cluster 1 for K-means ++ is a cluster of early middle age (age 34) Male, who is single. He graduates from High school and is currently Unemployed. He earns upper-middle income of $103,257 in a small city.
* Cluster 0 for Agglomerative Clustering is a cluster of late middle aged (age 47) Female, who is not single. She graduates from University and is a skilled employee. She is living in a big city and earns high-income, average at $168,085.
* Cluster 1 for Agglomerative Clustering is a cluster of early middle age (age 32) Male, who is single. He graduates from High school and is currently Unemployed. He earns upper-middle income of $95,041 in a small city.
* 837 customers identified as Cluster 1 by the Agglomerative Clustering method are 100% identified as Cluster 1 by the K-means ++ method as well. On the other hand, of the 1,163 customers identified as Cluster 0 by the Agglomerative Clustering method, only 976 of them are identified as Cluster 0 by the K-means ++ method. 187 of the customers identified as Cluster 0 by the Agglomerative Clustering method are classified as Cluster 1 by the K-means ++ method.

A close-up of a computer screen

Description automatically generated

**5. Recommendation**

* With Cluster 0, as she earns High Income and is living in big cities, promotes luxury retreats with natural surroundings. Emphasize luxuriousness, uniqueness of the trip, combined with nature aspect of the trip. Email newsletters weekly to keep her update with all the promotions of trips that are relevant to her liking. Design activities for kids/ elderly people as well because she most likely will bring her whole family. Combine the trip with luxury airline companies to create a all-in-one luxury bundle trip.
* With Cluster 1, as he is unemployed but has upper-middle income, chances are that he just left his job. He is looking for a trip as a relief at a budget. Promote group travel that encourages interaction and adventure to affordable places since he is single and unemployed. Leverage digital ads and partner with influencers who are resonate with his demographic to showcase travel experiences. Emphasize the budgeable of the trip and the promising of new adventures await. Create travel challenges or contests that encourage engagement and offer rewards in return.

**6. Conclusion**

* In conclusion, the report has divided the customer dataset into 2 clusters based on their characteristic using K-means ++ and Agglomerative Clustering. The 2 clusters from both methods are very similar to each other. Cluster 0 is female, non-single, graduates from University, employed and lives in a big city. She is middle-aged and earns a high income. Cluster 1 is male, single, graduates from High school, unemployed and lives in a small city. He is early middle-aged and earns upper-middle income. For cluster 0, emphasize on the luxury of the trip and its uniqueness, while for cluster 1, emphasize on the budgeable and adventurous of the trip.