Name: Aryan Kakran

PROJECT REPORT

Steps Covered:- 1,2,3,4

STEP 1: Deciding (not) to Segment

1: Implications of Committing to Market Segmentation :-

1.1 Commitment

- Market segmentation can be a powerful strategy, but it's not always the best choice. Before committing, organizations should consider the implications and long-term commitment involved.
- Market segmentation is a long-term strategy, requiring the organization's full commitment and willingness to make substantial changes and investments.

1.2 Cost Involvement

- Segmenting a market is not free. There are costs of performing the research, fielding surveys, and focus groups, designing multiple packages, and designing multiple advertisements and communication messages.
- Cahill advises segmentation only if the expected increase in sales exceeds the costs of implementing it.

1.3 Required Changes

- Changes may include developing new products, modifying existing ones, adjusting pricing, and altering distribution channels.
- The company's internal setup may also need to shift from focusing on products to focusing on market segments.

1.4 Executive -Level Decisions

- The decision to pursue market segmentation must be made at the highest executive level, with consistent communication and reinforcement throughout the organization.

2: Implementation Barriers :-

2.1 Senior Management Involvement

- Success depends on active leadership, resource allocation, and long-term support from senior management.

2.2 Organizational Culture

- Factors like resistance to change, poor communication, lack of consumer focus, and office politics can hinder segmentation.

2.3 Lack of Training

- Lack of knowledge about segmentation basics among leaders or team members can cause failure.

2.4 Objective restrictions

- Objective restrictions faced by the organisation, including lack of financial resources, or the inability to make the structural changes required.

2.5 Process Clarity

- Poor planning, unclear objectives, and lack of structured processes create hurdles.

3: Checklist

- This first checklist includes not only tasks, but also a series of questions which, if not answered in the affirmative, serve as knock-out criteria. For example: if an organisation is not market-oriented, even the finest of market segmentation analyses cannot be successfully implemented.

STEP 2: Specifying the Ideal Target Segment

Segment Evaluation Criteria

1 User Involvement

- User input is essential throughout the entire market segmentation analysis, not just at the beginning or end stages.

2 Organizational Contribution

- After deciding to explore segmentation, the organization plays a key role by guiding key decisions, like data collection and choosing target segments.

3 Two Types of Criteria

3.1 Knock-Out Criteria

- These criteria are the essential, non-negotiable features for market segments to qualify as potential target segments. They act as initial filters, and any segment that fails to meet them is eliminated from further consideration.
  Kotler proposed key criteria like:
  - **Homogeneous:** Members within a segment should be similar.
  - **Distinct:** Segments should be clearly different from each other.
  - **Large Enough:** The segment should be big enough to justify extra marketing efforts.
  - **Matching Strengths:** The organization should be capable of meeting the segment's needs.
  - **Identifiable:** Segment members should be easy to identify in the market.
  - **Reachable:** There should be a way to connect with the segment effectively.

3.2 Attractiveness Criteria

- Attractiveness criteria rate how appealing each segment is, rather than just being a yes or no decision. These ratings guide whether a segment becomes the target in the final step of market segmentation.

4 Criteria from Literature

- Different researchers propose various evaluation criteria such as measurability, size, profitability, and accessibility. These criteria help assess segments' viability and alignment with organizational goals.

5 Applying Criteria

- Knock-out criteria quickly eliminate segments that do not fit.
- The attractiveness criteria are then used to rank the remaining segments based on their potential.

6 Implementing a Structured Process

Experts agree that using a structured process to evaluate market segments is beneficial. A popular method is the segment evaluation plot, which compares segment attractiveness with organizational competitiveness. The team selects these factors, focusing on the most relevant ones for the organization.

- Choose up to six important factors for attractiveness and competitiveness.
- Include representatives from different organizational units for their perspectives.
- Agree on attractiveness criteria early to guide data collection and make segment selection easier later.

STEP 3: Collecting Data

Segmentation Variables:

- Empirical data forms the basis for both commonsense and data-driven segmentation. In commonsense segmentation, a single characteristic (segmentation variable) is used to split a sample into segments. Other characteristics serve as descriptor variables that describe the segments.

Descriptor Variables:

- Descriptor variables provide detailed descriptions of segments and include factors like age, vacation frequency, and benefits sought. This helps in developing a targeted marketing mix.

Commonsense vs. Data-Driven Segmentation:

- Commonsense segmentation relies on a single segmentation variable to split a market (e.g., gender).
- Data-driven segmentation uses multiple segmentation variables to identify or create segments.

Segmentation Criteria:

- Organizations choose a segmentation criterion before collecting data. The segmentation criterion defines the nature of information used for segmentation and may relate to a construct like benefits sought.

Geographic Segmentation:

- This approach uses consumers' residence location as the segmentation criterion, often proving simple and effective.

Socio-Demographic Segmentation:

- Criteria like age, gender, income, and education are used to segment markets based on demographic attributes, which are useful for industries like luxury goods or baby products.

Psychographic Segmentation:

- Grouping people based on psychological criteria (e.g., beliefs, interests, aspirations) is referred to as psychographic segmentation. Benefit segmentation and lifestyle segmentation are key examples.

Behavioural Segmentation:

- Segmentation based on observed or reported behaviors. Various consumer behaviors can serve as segmentation criteria.

Data from Survey Studies:

Market segmentation often relies on survey data. Important considerations include:

- Choice of Variables: Carefully selecting variables in segmentation.
- Response Options: Answer choices determine the data scale.
- Response Styles: Surveys may capture biases affecting segmentation.
- Sample Size: There are challenges regarding sample size for segmentation algorithms.

Data from Internal Sources:

- Organizations often use internal data to support segmentation analysis.

Data from Experimental Studies:

- Experimental data from field or lab experiments, like responses to advertisements, can be used in segmentation.

STEP 4: Exploring Data

Exploratory Data Analysis (EDA) is the first step in working with a dataset. It helps us understand the data better before we analyze it further. This stage is crucial for understanding the data's structure and determining the best algorithms for extracting insights, like market segments.

The technical tasks involved in data exploration include:

1. Identifying the measurement levels of variables.
2. Investigating the univariate distributions of each variable.
3. Assessing the dependency structures among variables.

Reading Data from a CSV File

To read data from a CSV file, use the following code:

```
vacation_df = pd.read_csv("vacation.csv")
```

Columns in dataset
To display the columns present in the dataset:
```
vacation_df.columns
```

Shape of dataset
To check the shape (number of rows and columns) of the dataset:
```
vacation_df.shape
```

Description of dataset
To get a descriptive summary of the dataset, including all columns:
```
description = vacation_df.describe(include='all')
print(description)
```

DATA CLEANING
Data cleaning is the process of identifying and correcting errors or inconsistencies in data to improve its quality and reliability. It's a crucial step in data analysis, ensuring that the data you work with is accurate, complete, and ready for meaningful analysis.
This involves:

1. **Check for Errors**: Ensure all data is recorded correctly and that categories have consistent names.
2. **Validate Values**: Make sure numerical values fall within expected ranges and that categorical data only includes valid options.
3. **Manage Levels**: For categorical data, check that the levels (or categories) are correct and in the right order.

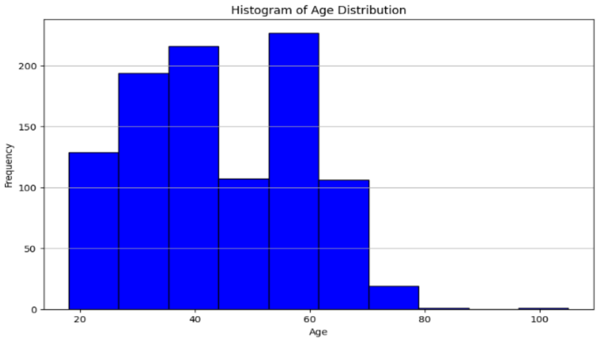
```
inc2 = vacation_df['Income2']

original_levels = inc2.unique()

new_levels = ["<30k", "30-60k", "60-90k", "90-120k", ">120k"]


inc2_new = pd.Categorical(inc2, categories=new_levels, ordered=True)
vacation_df['Income2'] = inc2_new

print(vacation_df['Income2'].cat.categories)
print(vacation_df['Income2'].value_counts())
```

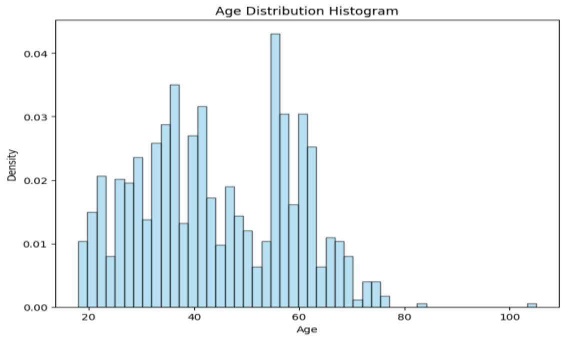
DESCRIPTIVE ANALYSIS
Descriptive Analysis provides insights and avoids misinterpretation of complex data.

Graphical Methods for Numeric Data:

- **Histograms**: Visualize data distribution; show how often values occur within specific ranges.

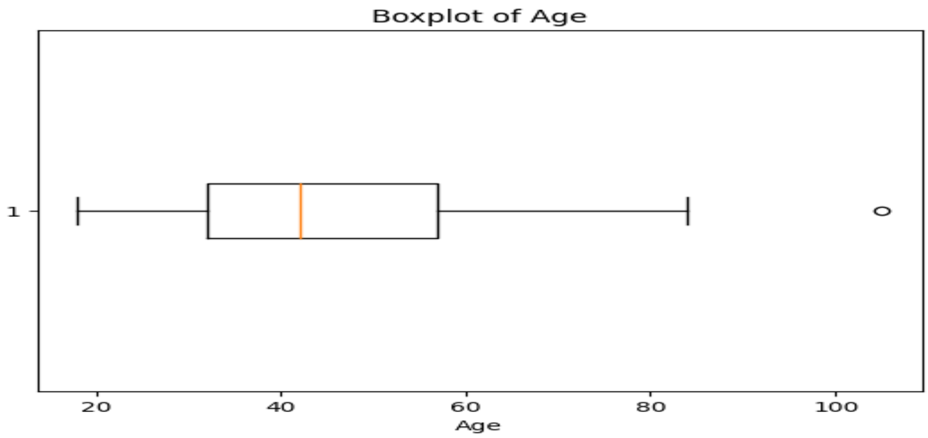
```
plt.figure(figsize=(10, 6))
plt.hist(vacation_df['Age'], bins=10, color='blue', edgecolor='black')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Histogram of Age Distribution')
plt.grid(axis='y')
plt.show()
```





- **Boxplots**: Display key statistics (min, quartiles, median, max) and identify outliers.

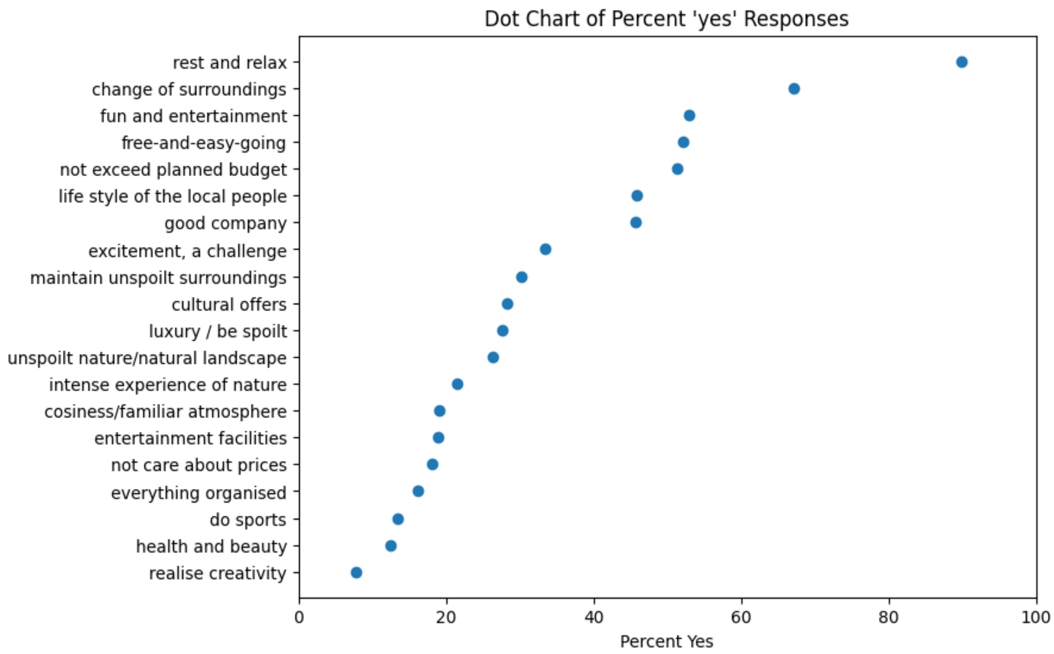
```
plt.boxplot(vacation_df['Age'], vert=False)
plt.xlabel('Age')
plt.title('Boxplot of Age')
plt.show()
```



- **Scatter Plots**: Show relationships between numeric variables.

```
plt.figure(figsize = (8,6))
plt.scatter(yes_sorted,yes_sorted.index)
plt.xlabel("Percent Yes")
```

```
plt.xlim(0,100)
plt.title("Dot Chart of Percent 'yes' Responses")
plt.show()
```



6.4 Pre-Processing

6.4.1 Categorical Variables

- **Merging Levels**: Useful for reducing categories when there are too many differentiated levels.
    - Example: Income categories showed many respondents in lower ranges; merging resulted in a more balanced distribution.
- **Conversion to Numeric**:
    - Ordinal data (e.g., income) can be converted to numeric if the distances between categories are approximately equal.
    - Likert scales, commonly used in surveys, may not have equal distances; therefore, binary options might be preferable to avoid response style biases.
- **Binary Variables**:
    - Easier to process and less affected by response styles. Can be easily converted to numeric (0/1).
    - Example conversion from travel motives to a numeric matrix for segmentation analysis.

```
vacmot = (vacation_df == "yes").astype(int)
```

6.4.2 Numeric Variables

- **Standardization**: Necessary to balance the influence of variables with different scales.
  - Standardization formula: zi=xi−x¯sz_i = \frac{x_i - \bar{x}}{s}zi=sxi−x¯, where x¯\bar{x}x¯ is the mean and sss is the standard deviation.
  - Standardization can be done in R using scale().

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
vacmot_scaled=scaler.fit_transform(vacmot)
vacmot_scaled
```

- **Outliers Consideration**: Robust methods (median and interquartile range) may be required for data with outliers.

Principal Components Analysis (PCA)

1. **Purpose of PCA:**
   - Transforms multivariate data into principal components that are uncorrelated and ranked by importance.
2. **Components:**
   - The first principal component captures the most variability; subsequent components capture less.
3. **Data Preservation:**
   - The relative positions of observations remain unchanged, and the number of variables remains the same.
4. **Working Mechanism:**
   - Utilizes the covariance or correlation matrix of numeric variables.
   - If variables differ in scale, use the correlation matrix.
5. **Dimensionality Reduction:**
   - PCA is often used to reduce high-dimensional data to lower dimensions for visualization, typically using the first two or three principal components.
6. **Variance Explained:**
   - Summarizes the proportion of variance explained by each component, showing that the first few components often explain significant variation, while later ones explain much less.
7. **Visualization:**

- Two-dimensional plots can be created using the first two principal components, allowing for easier interpretation of the data structure.

8. **Unique Insights:**

- Components can reveal distinctive travel motives; for example, certain components may highlight specific preferences (like budget constraints).

Implementation code:

```
from sklearn.decomposition import PCA
pca = PCA()

vacmot_pca = pca.fit_transform(vacmot)
vacmot_pca

print("Explained Variance (importance of each PC):", pca.explained_variance_)

print(pd.DataFrame(pca.components_.T, index=vacmot.columns, columns=[f'PC{i+1}' for i in range(len(vacmot.columns))]))


pca.fit(vacmot)

# Create a summary dataframe for explained variance
summary_df = pd.DataFrame({
    'Component': [f'PCA{i+1}' for i in range(len(pca.explained_variance_ratio_))],
    'Standard Deviation': np.sqrt(pca.explained_variance_),
    'Proportion of Variance': pca.explained_variance_ratio_,
    'Cumulative Proportion': np.cumsum(pca.explained_variance_ratio_)
})

summary_df = summary_df.round(2)
print(summary_df)




#PCA Projection
projected_data = pca.fit_transform(vacmot)
feature_names = vacation_df.columns.to_list()

# Plot the second and third components
plt.figure(figsize=(15, 11))
plt.scatter(projected_data[:, 1], projected_data[:, 2], color='white', alpha=0.5, edgecolor='k', s=30)
```

```
plt.xlabel('PC2')
plt.ylabel('PC3')

for i, feature in enumerate(feature_names):
    arrow_x = pca.components_[1, i] * max(projected_data[:, 1])
    arrow_y = pca.components_[2, i] * max(projected_data[:, 2])

    # Plot the arrow
    plt.arrow(0, 0, arrow_x, arrow_y, color='red', width=0.01, head_width=0.05,
head_length=0.1, length_includes_head=True)

    plt.text(arrow_x * 1.1, arrow_y * 1.1, feature, color='blue', fontsize=10)

plt.title('Projection on PC2 and PC3 with Feature Axes')
plt.grid(True)
plt.show()
```

Github Repository link:
https://github.com/akakran22/Project2.git

Profile :-
https://github.com/akakran22