**MAVEN MARKET CUSTOMERS DATA ANALYSIS DOCUMENTATION**

**OUTLINE**

**1. Introduction**

**2. Data Cleaning**

**3. Analysis Sheet**

**4. Visualization**

**5. Dashboard Creation**

**6. Challenges Faced**

**7. Insights from the Dataset**

**8. Recommendations**

**INTRODUCTION**

The project revolves around comprehensive analysis of customer data sourced from the MavenMarket dataset. Specifically, it focuses on the MavenMarket\_Customers CSV file, which is part of a collection of datasets including:

* MavenMarket\_Calender,
* MavenMarket\_Products,
* MavenMarket\_Regions,
* MavenMarket\_Returns\_1997-1998,
* MavenMarket\_Stores, and
* MavenMarket\_Transactions\_1998.

The MavenMarket\_Customers dataset comprises various fields including customer\_id, acct\_num, first\_name, last\_name, address, city, state\_province, postal\_code, country, birthdate, marital\_status, yearly\_income, gender, total\_children, num\_children\_at\_home, education, acct\_open\_date, member\_card, occupation, and homeowner.

The primary objective of the project is to extract actionable insights from the customer data through meticulous data cleaning, analysis, visualization, and dashboard creation. By delving into customer demographics, behavior, and preferences, the project aims to inform strategic decision-making processes for the MavenMarket organization.

Through the utilization of advanced Excel functionalities, including pivot tables, charts, and slicers, the project endeavors to provide stakeholders with intuitive and visually appealing representations of the data. These representations facilitate easy interpretation of complex information, enabling stakeholders to identify trends, patterns, and opportunities within the customer dataset.

The project is structured to ensure transparency, efficiency, and usability, with each phase meticulously documented to provide a comprehensive overview of the methodologies employed and the insights derived. Through this approach, stakeholders and users are empowered to make informed decisions and drive actionable outcomes based on the findings of the analysis.

**DATA CLEANING**

Data cleaning is a crucial initial step in any data analysis project, aimed at ensuring data accuracy, consistency, and completeness. In the context of this project, the MavenMarket\_Customers dataset underwent meticulous data cleaning procedures, which included the following steps:

1. **Checking for Blanks**: Each field in the dataset was examined to identify any missing or blank values. This process helps ensure data integrity and completeness. Fortunately, no blanks were found in any field, indicating a high level of data quality.

2. **Organizing Data into a Table**: To facilitate easier management and analysis of the dataset, the raw data was structured into a tabular format. This allows for easier manipulation and application of Excel functionalities such as pivot tables and charts.

3. **Adjusting Data Types:** The next step involved reviewing and adjusting the data types of each field to ensure compatibility with the intended analysis. Excel's built-in functionalities were utilized to convert data types to appropriate formats, such as dates, text, or numerical values. This step helps prevent errors and inconsistencies during subsequent analysis.

4. **Value Replacements:** Certain values within specific fields were replaced to enhance clarity and consistency across the dataset. For example:

- Marital\_status field: 'M' was replaced with 'Married', and 'S' with 'Single'.

- Gender field: 'F' was replaced with 'Female', and 'M' with 'Male'.

- Homeowner field: 'Y' was replaced with 'Yes', and 'N' with 'No'.

By implementing these data cleaning procedures, the MavenMarket\_Customers dataset was prepared for further analysis. The integrity and quality of the data were ensured, laying a solid foundation for extracting meaningful insights and trends in the subsequent phases of the project.

**ANALYSIS SHEET**

The "Analysis Sheet" serves as the central hub for conducting in-depth analysis of the MavenMarket\_Customers dataset. This sheet houses a series of pivot tables that are strategically designed to extract valuable insights into various aspects of customer demographics, behavior, and preferences.

Each pivot table on the Analysis Sheet serves as a powerful tool for extracting actionable insights from the MavenMarket\_Customers dataset. Together, they provide a comprehensive understanding of various dimensions of customer behavior and characteristics, laying the groundwork for informed decision-making and strategic planning.

Below is a breakdown of the pivot tables created on the Analysis Sheet:

1. **Distinct Count of Customer Attributes:**

* Pivot tables were created to calculate the distinct count of several customer attributes, including:
* customer\_id: Provides insights into the total number of unique customers in the dataset.
* city: Indicates the number of unique cities represented in the dataset.
* state\_province: Reveals the diversity of states or provinces where customers are located.
* member\_card: Illustrates the variety of member card types held by customers.
* country: Shows the distribution of customers across different countries.

2. **Customer Account Creation by Month:**

* This pivot table tracks the monthly account creation trends of customers. It enables the visualization of customer acquisition patterns over time, providing valuable insights into seasonal variations or trends in account openings.

3. **Count of Customers by Country:**

* By aggregating the count of customers based on their respective countries, this pivot table offers a comprehensive overview of the geographical distribution of customers.

4. **Customer Member Card Type by Customer Education:**

* This pivot table examines the relationship between customer education levels and their member card types. It allows for the identification of any patterns or correlations between education and membership preferences.

5. **Percentage Distribution of Customer Gender and Marital Status:**

* These pivot tables calculate the percentage distribution of customer gender and marital status categories. They provide insights into the gender and marital composition of the customer base, facilitating targeted marketing strategies or product offerings.

6. **Percentage Distribution of Registered Member Card Types:**

* By analyzing the percentage distribution of registered member card types, this pivot table helps identify the popularity of different membership programs among customers.

7. **Top 5 Locations by Country:**

* This pivot table identifies the top five locations (state\_province and city) within each country based on the maximum number of customers. It highlights key geographic areas with the highest customer concentrations.

8. **Member Card Population by Yearly Income Rate:**

* By categorizing customers based on their yearly income rates, this pivot table provides insights into the distribution of member card populations across income brackets.

9. **Customers' Total Number of Children Count:**

* This pivot table calculates the total number of children count for customers, enabling the analysis of family demographics and potential market segments.

**VISUALIZATION**

In this project, the analysis conducted on the "Analysis Sheet" was visually represented through charts, ensuring that stakeholders can grasp key findings at a glance.

The carefully curated visualizations on the dashboard provide a user-friendly interface for exploring and understanding the wealth of information contained within the MavenMarket\_Customers dataset.

Here's how the visualization phase was executed:

1. **Creation of Charts:**

* Charts were generated based on the pivot tables created on the Analysis Sheet. Each chart was carefully chosen to effectively represent the insights derived from the analysis.
* Charts types such as bar charts, pie charts, and line graphs were selected based on the nature of the data being visualized and the insights being conveyed.
* The data presented in the pivot tables, such as distinct counts, percentages, and trends, were translated into visual formats for easier interpretation.

2. **Chart Adjustments:**

* The charts underwent meticulous adjustments to enhance clarity and visual appeal. This included:
* Removal of unnecessary features such as gridlines and data labels where not needed, ensuring a clean and uncluttered appearance.
* Hiding all field buttons to streamline the appearance of the charts and minimize distractions.
* Adding and adjusting data labels to provide context and aid interpretation of the data points.
* Editing chart titles to clearly communicate the insights conveyed by the visualizations.
* Changing chart styles and colors to ensure consistency with the overall aesthetic and improve readability.

3. **Chart Integration:**

* Once the charts were refined, they were seamlessly integrated into the dashboard on the "Dashboard" sheet. This ensured that all visualizations were consolidated in one central location for easy access and reference.
* The visualizations were strategically arranged on the dashboard to provide a cohesive and intuitive overview of the key insights derived from the analysis.

4. **Purposeful Visualization:**

* Each visualization served a specific purpose in communicating key findings and trends to stakeholders. Whether highlighting customer demographics, geographic distribution, or membership preferences, the visualizations were designed to convey actionable insights that inform decision-making processes.

**DASHBOARD CREATION**

The "DASHBOARD" sheet serves as the central interface for stakeholders to interact with and derive insights from the visualizations and analyses conducted on the MavenMarket\_Customers dataset.

By combining visually appealing design elements with interactive features such as slicers and summary metrics, the dashboard provides stakeholders with a user-friendly and informative interface for exploring and understanding the insights derived from the MavenMarket\_Customers dataset.

Here's a detailed overview of the dashboard creation process:

1. **Visual Enhancement:**

* The dashboard was designed with a focus on visual appeal and usability. A background was added to enhance the overall aesthetic appeal of the dashboard.
* Reports and visualizations were meticulously aligned and arranged to ensure a clean and organized layout, facilitating ease of access and navigation for users.

2. **Key Summary Metrics:**

* At the top of the dashboard, key summary metrics were prominently displayed in the form of summary cards. These cards provide quick and concise insights into important aspects of the dataset, including distinct counts of customers, states, cities, countries, and member card types. This allows stakeholders to gain immediate understanding of essential metrics without delving into detailed analyses.

3. **Labeling and Descriptive Titles:**

* Each visualization on the dashboard was labeled and provided with descriptive titles to clearly communicate the insights conveyed by the visualizations. This ensures that users can quickly interpret the meaning of each visualization without ambiguity.

4. **Removal of Redundant Elements:**

* Redundant elements and features that could potentially clutter the dashboard were removed to maintain a clean and streamlined appearance. This includes unnecessary gridlines, data labels, or other elements that do not contribute to the visualization's clarity or usability.

5. **Slicer Integration:**

* Slicers were added vertically on the left side of the dashboard to provide users with dynamic filtering options. Slicers for gender, marital\_status, country, state, and member\_card enable users to interactively explore the data and customize the visualizations based on specific criteria. This enhances user engagement and allows for deeper insights into different segments of the customer dataset.

6. **Color Format and Dashboard Heading:**

* A simple color format was employed throughout the dashboard to maintain clarity and consistency in visual presentation. Bold descriptive headings were placed prominently at the top of the dashboard, ensuring users can easily identify the purpose and scope of the dashboard.

7. **Visual Locking and Editing Restrictions:**

* To preserve the integrity of the dashboard layout and prevent unintended changes, visuals were locked, and moving and resizing were disabled for necessary reports. This ensures that the dashboard maintains its intended design and functionality while still allowing users to interact freely with the visualizations and slicers.

**CHALLENGES FACED**

Despite the successful execution of the project, several challenges were encountered throughout the various phases. These challenges, though initially disruptive, were effectively addressed, contributing to the overall success of the project. Here are the key challenges faced:

**Pivot Table Connections:**

One significant challenge arose from the initial failure to add pivot tables to the data model. This oversight resulted in slicers being unable to connect to most pivot tables and fields unable to utilize the 'distinct count' function. This hindered interactivity and limited the functionality of the dashboard.

**Data Model Integration:**

The absence of pivot tables from the data model necessitated a retrofitting process to rectify the issue. Integrating pivot tables into the data model was crucial to enable slicers to function correctly and unlock the 'distinct count' functionality. This required additional time and effort to troubleshoot and implement.

**Data Quality Assurance:**

While no blanks or missing values were detected during the initial data cleaning phase, ensuring data quality and accuracy remained a priority throughout the project. Continuous validation and verification processes were implemented to identify and rectify any inconsistencies or errors that may have arisen during subsequent phases.

**Dashboard Usability Testing:**

Ensuring the usability and functionality of the dashboard posed another challenge. Testing the dashboard's interactivity, responsiveness, and ease of navigation was essential to guarantee a seamless user experience. Iterative testing and refinement were conducted to address any usability issues and optimize user engagement.

**Resource Constraints:**

Limited resources, such as electricity and low processing speed of PC, presented challenges in meeting set project deadlines and implementing complex functionalities. Prioritization of tasks and efficient utilization of available resources were critical to overcoming these constraints and delivering a successful project outcome.

**Supervisor Communication:**

Effective communication with my supervisors throughout the project lifecycle was essential for managing expectations, gathering requirements, and addressing feedback.

**INSIGHTS FROM THE DATASET**

These insights provide valuable information about the demographics, preferences, and behavior of customers in the MavenMarket dataset, enabling informed decision-making and targeted strategies to enhance customer satisfaction and retention.

**Distinct Count of customer\_id:**

The dataset contains a total of 10,281 unique customer IDs, indicating a diverse customer base.

**Distinct Count of city:**

There are 108 unique cities represented in the dataset, highlighting the geographic spread of customers.

**Distinct Count of state\_province:**

A total of 13 states or provinces are represented in the dataset, suggesting a regional distribution of customers.

**Distinct Count of member\_card:**

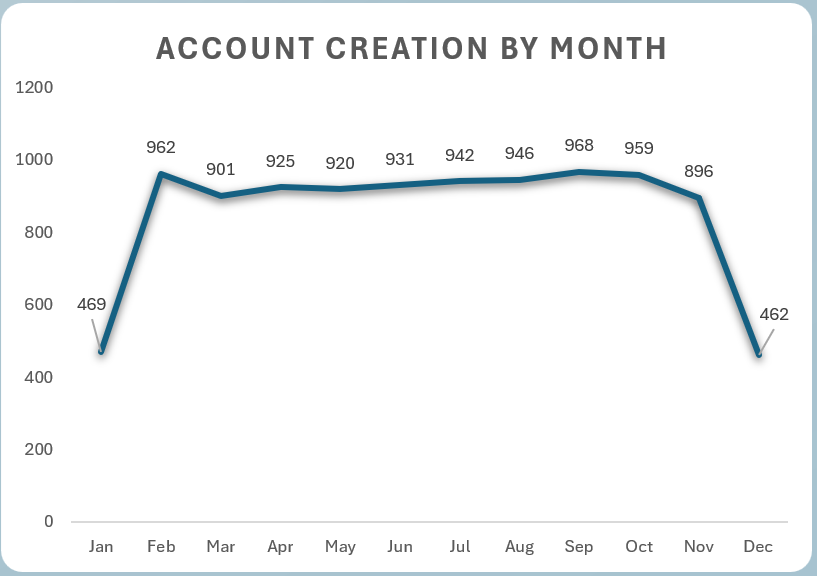
The dataset includes four types of member cards: bronze, golden, normal, and silver.

**Distinct Count of country:**

Customers are primarily from three countries: USA, Canada, and Mexico.

**Account Creation by Month:**

Account creation exhibits some variability across months, with February having the highest count (962) and December the lowest (462).



**Count of Customers by Country:**

USA has the highest number of customers (7,359), followed by Canada (1,717), with Mexico having the fewest customers.

A graph with numbers and a bar

Description automatically generated

**Top Locations by Country:**

The states with the most customers include Shawnee in BC Canada, Downey in CA USA, Lebanon in OR USA, Lemon Grove in CA USA, and Richmond in CA and BC USA and Canada respectively.



**Member Card Usage by Education:**

Bronze card usage is highest among customers with high school and BSc degrees, while golden card usage is predominantly by high school degree holders. Normal card usage is spread across various education levels, with fewer users outside of partial high school degrees. Silver card usage is highest among high school degree holders.

A graph of different colored bars

Description automatically generated

**Member Card Usage Percentage:**

Bronze card usage constitutes 55.47% of total member card usage, followed by normal card usage (23.54%), golden card usage (11.65%), and silver card usage (9.34%).

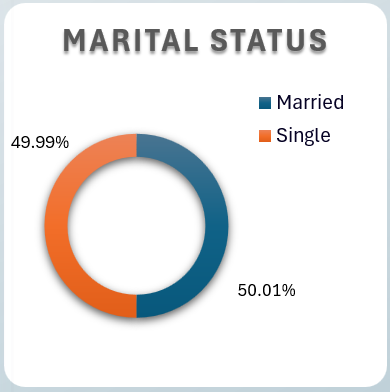
A close-up of a member card

Description automatically generated

**Gender and Marital Status Distribution:**

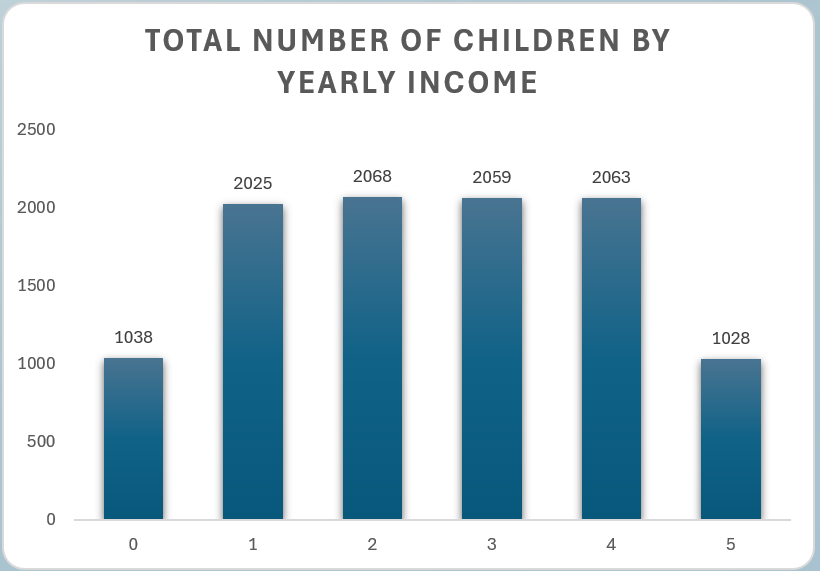
Male customers slightly outnumber female customers. There is a minimal difference in the number of single and married customers.

A graph of gender

Description automatically generated with medium confidence

**Total Number of Children by Yearly Income:**

Most customers have four children, with relatively even distribution among customers with 0 to 5 children.



**Member Card Population by Yearly Income:**

The distribution of member card types varies across different income brackets, with bronze and normal cards being more prevalent across all income ranges.

A graph of income from a member card

Description automatically generated

**RECOMMENDATIONS**

Having derived valuable insights from the MavenMarket dataset, it is essential to outline the next steps to leverage these findings effectively and drive actionable outcomes. Here are the recommended next steps:

* Utilize the insights gained to segment customers based on their characteristics and preferences. Develop targeted marketing strategies and personalized offerings tailored to different customer segments to enhance engagement and satisfaction.
* Implement retention strategies, such as loyalty programs or personalized communications, to strengthen customer relationships and improve retention rates.
* Identify areas of opportunity for introducing new products or enhancing existing ones to better meet customer needs and preferences.
* Implement mechanisms for collecting feedback from customers to gain deeper insights into their preferences, satisfaction levels, and pain points. Leverage customer feedback to iterate on strategies and enhance the overall customer experience.
* Embrace a culture of continuous learning and improvement by staying abreast of industry trends, emerging technologies, and best practices in customer analytics. Invest in training and development opportunities for team members to enhance their skills and capabilities in data analysis and interpretation.