**Bank Market Campaign Project**

**First step:**

**Title:** Converting Semicolon-Delimited Data to Table-Formatted Excel

**Summary:** This code snippet demonstrates how to convert semicolon-delimited data from a CSV file into a table-formatted Excel file using the pandas library in Python. It also includes additional steps for data preprocessing and analysis.

Code: Python Code

A screenshot of a computer program

Description automatically generated

**Data Cleaning and Data processing**

**Duplicated Data:**

In Excel, 14 duplicated records were identified and removed using the "Remove Duplicates" function. This step ensures that the dataset contains only unique entries, preventing any redundancy.

**Marital Unknown Values:** Unknown marital status values were replaced with the most frequent marital status type. This process enhances data consistency and enables more meaningful analysis.

**Formula:** =index(c:c,Mode(if(c:c)",match(c:c,c:c,0))))

**Job Category:** A new column, "Job\_Category," was created based on the "Job" column. Customers were categorized into different job categories, including "Working Class," "Professional Class," "Non-Working," and "Unknown." This categorization simplifies job-related analysis.

**Formula:** =IF(OR(B2="housemaid", B2="blue-collar", B2="technician", B2="admin.", B2="student"), "Working Class",

IF(OR(B2="entrepreneur", B2="management", B2="services", B2="self-employed"), "Professional Class",

IF(OR(B2="unemployed", B2="retired"), "Non-Working",IF(B2="unknown", "Unknown", B2))))

**Education Category:** Another new column, "Education\_Category," was generated from the "Education" column. Customers' education levels were grouped into categories such as "Low-Education," "High-School-Education," "Professional-Education," and "Unknown." This categorization facilitates education-related insights.

**Formula**: =IF(OR(F2="basic.4y", F2="basic.6y", F2="basic.9y", F2="illiterate"), "Low-Education",

IF(F2="high.school", "High-School-Education",

IF(OR(F2="professional.course", F2="university.degree"), "Professional-Education",

IF(F2="unknown", "Unknown", ""))))

**Duration Category:** A new column, "Duration\_Category," was introduced using data from the "Duration" column. Customers' contact durations were categorized as "Short\_Duration," "Medium\_Duration," or "Long\_Duration." This categorization aids in analyzing the impact of contact duration on outcomes.

**Formula:** =IF(N2<=1000, IF(N2<=500, "Short\_Duration", "Medium\_Duration"), "Long\_Duration")

**Campaign Category:** The "Compaign" column was used to create a new column, "Campaign\_Category," which classifies campaigns based on contact frequency. Campaigns were categorized as having "Low\_Contact\_Frequency," "Medium\_Contact\_Frequency," or "High\_Contact\_Frequency." This categorization assists in evaluating the effectiveness of different campaign types.

**Formula:** =IF(AND(P2>=1, P2<=5), "Low\_Contact\_Frequency", IF(AND(P2>=6, P2<=10), "Medium\_Contact\_Frequency", "High\_Contact\_Frequency"))

**Data Analysis to answer some business questions**

**Business Question 1**: How does the distribution of education levels vary among different job categories, and are there any insights we can derive from this?

**Answer:**

Working-Class: The "Working-Class" job category exhibits a diverse distribution of education levels. The highest count is observed in the "Low-Education" category (9,415), followed by "Professional-Education" (12,197). This indicates that a significant number of working-class individuals have education levels below the professional category.

Professional-Class: In the "Professional-Class" job category, there's an expected presence of customers with "Professional-Education" (4,221). However, intriguingly, a significant number of customers (1,858) also fall into the "Low-Education" category. This suggests a diversity in education levels within the professional job category.

Non-Working: Among "Non-Working" individuals, "Low-Education" (1,152) is the most common education level, followed by "Professional-Education" (929). This implies that a substantial portion of non-working individuals in the dataset has lower education levels.

Unknown: The "Unknown" job category exhibits lower counts across various education levels, indicating less available information for this category. Further investigation is needed to understand this group better.

Overall, this analysis highlights the distribution of education levels across job categories. While certain patterns emerge, it's important to consider additional factors that may influence these distributions. Further analysis and context are necessary to draw actionable insights.

Visual Representation: The horizontal bar chart visually represents this distribution. Each bar corresponds to a job category, and the bar length indicates the count of customers in each education category within that job category. This chart facilitates a quick comparison of education distributions across job categories.

**Business Question 2:** How do conversion rates for "no" and "yes" responses vary across different months, and are there specific months with a higher likelihood of positive responses?

**Answer:**

The table displays the percentage distribution of "no" and "yes" responses across different months. Notably, May stands out with a relatively high "yes" response rate (19.10%), indicating a higher likelihood of positive responses during this month. Conversely, December has the lowest "yes" response rate (1.92%). Seasonal patterns emerge, with May being a strong month for positive responses, while other months like December and March have lower conversion rates.

The grand total percentages reveal that the dataset maintains a reasonable balance between "no" and "yes" responses. Additionally, April exhibits a relatively high "yes" response rate (11.62%), suggesting it may be a favorable month for marketing campaigns.

**Business Question 3:** Is there a correlation between the customer's age (x-axis) and the number of times they were previously contacted in previous marketing campaigns (y-axis), and how might this correlation influence future campaign strategies?

**Answer:**

Based on the scatter plot, it's challenging to discern a clear correlation between customer age (x-axis) and the number of previous marketing campaign contacts (y-axis). However, the plot suggests that more customers have been contacted fewer times compared to those contacted more times.

To interpret the correlation's influence on future campaigns, additional data and analysis are needed. One potential strategy could involve targeting customers who have been contacted fewer times, as they might be more receptive to future marketing efforts. These explanations provide a clear and meaningful understanding of the analyses and their implications for your Excel project. They emphasize the insights gained and their relevance to decision-making in the context of the dataset.