**Exploration of Customer Churn Data**

Kelseyann Wright

Student ID: 00565814

College of Information Technology, Western Governors University

Data Exploration - D207

Dr. William Sewell

March 22, 2023

**Exploration of Customer Churn Data**

**A1. Question for Analysis**

The question that will be addressed through this data exploration is, “Is the type of contract related to customer churn?”

**A2. Benefit from Analysis**

Stakeholders in the organization will benefit from this analysis because it will help identify or eliminate a potential factor related to customer churn. If customers with shorter contracts are more or less likely to discontinue service, that information will be beneficial for marketing and sales strategies. Understanding how the length of a contract is related to customer churn is valuable information for customer retention long-term.

**A3. Data Identification**

The features in the churn dataset that will be relevant to this analysis are “churn” and “contract”. In addition, the relationships with income and monthly charge will be visualized and distributions identified (section C and D). The contract and churn variables are categorical. The income and monthly charge variables are continuous numeric. The final dataset to use for the analysis is 10000 rows and 4 columns. A chi-square test will be used to assess the relationship between contract type and customer churn because both are categorical variables (Bruce and Gedeck, 2020; Griffiths, 2008). The independent variable is contract type, and the dependent variable is customer churn.

**B1. Code**

The code is attached (D207.ipynb).

Text

Description automatically generated**B2. Output**

A chi-square test for independence was performed (Figure 1). The test statistic was 718.59, and the p-value was 9.1 x 10-157 (Figure 1).

**B3. Justification**

The chi-square test was performed using the SciPy package in python (Vertanen et al., 2020). The chi-square test enables the analyst to determine if two variables are related or dependent on each other. This directly answers our question, “Is the type of contract related to customer churn?”. The chi-square test is appropriate to assess the relationship between the variables “contract” and “churn” are both categorical variables. The chi-square test does not require a test for normality like the other methods.

**Figure 1: Chi square test output**

**C. Univariate Statistics**

Histograms were made to visualize distributions of the continuous variables (income and monthly charge, Figure 2). Bar plots of counts in each category were created to visualize distributions of the categorical variables of churn and contract (Figure 2). The income variable has a right (positive) skewed distribution (Figure 2, Bruce and Gedeck, 2020). The mean income is $39,806.92 with a maximum income of $258,900.70. Monthly charge appears platykurtic with a slight positive skew (Figure 2, Bruce and Gedeck, 2020). The mean monthly charge is $172.62. The churn variable contains two categories: yes and no. The dataset contains 7350 “no” churn customers and 2650 “yes” churn customers (Figure 2). The contract variable contains three categories: month to month, one year, and two years. There are 5456 month-to-month customers, which accounts for over half of the total customers (Figure 2). There are 2102 customers with one-year contracts, and 2442 customers with two-year contracts (Figure 2).

**C1. Visual of Findings**

Chart

Description automatically generated

Figure 2: Univariate graphs for determining distributions of variables.

**D. Bivariate Statistics**

The bar graph comparing customer churn across the three contract types shows less customer churn for customers with one year and two-year contracts than customers with month-to-month contracts (Figure 3). 37.3% of customers with month-to-month contracts churned, while only 14.6% of customers with one-year contracts and 12.7% of customers with two year contracts churned. Customer income was consistent across all contract types and churn categories. Monthly charge appears to be slightly higher for churn customers than customers that did not drop service.

**D1. Visual of Findings**

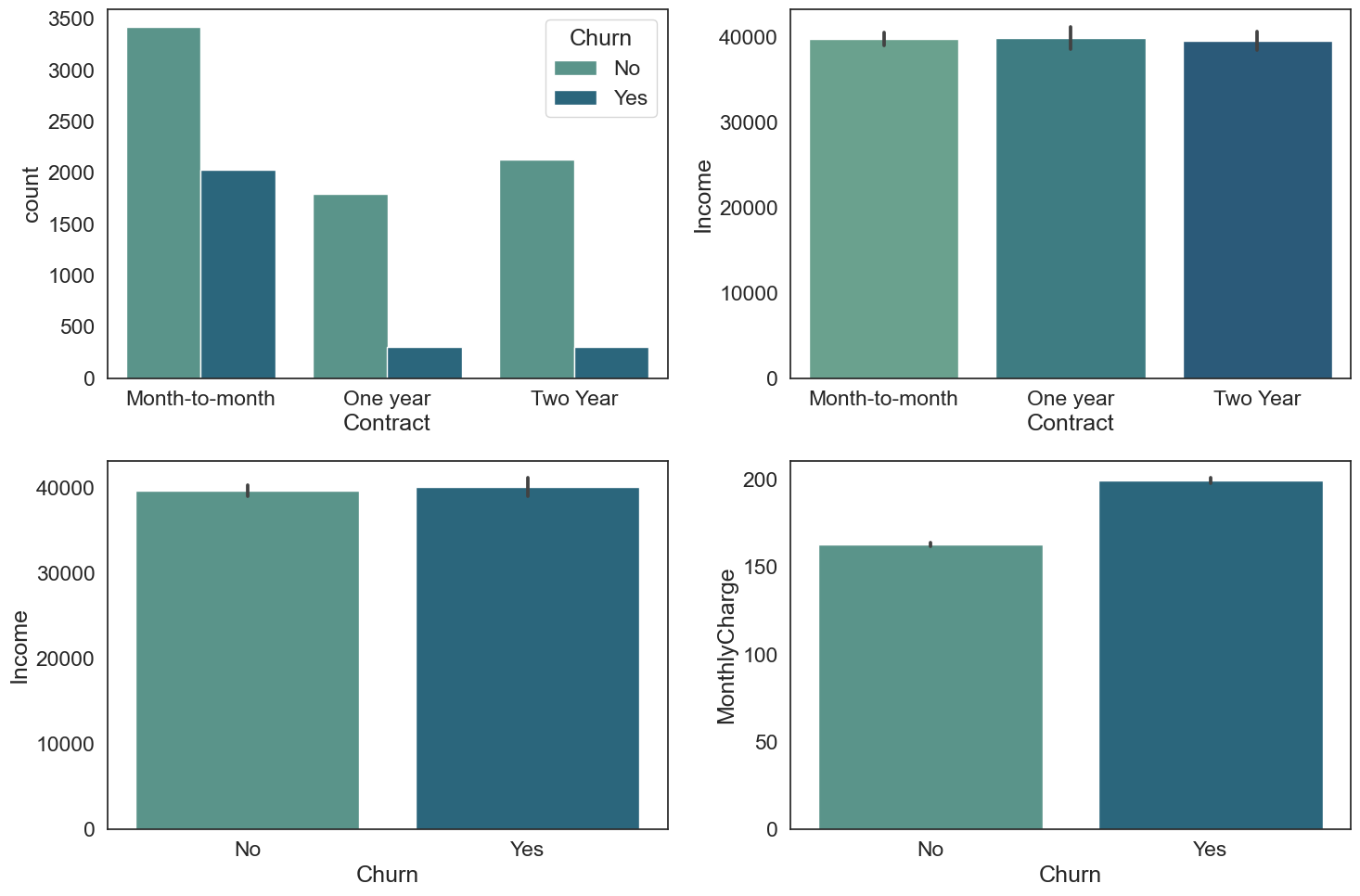
****

Figure 3: Bivariate graphs for determining distributions of variables.

**E1. Results of Analysis**

A chi-square test for independence was performed because both the independent variable (contract type) and dependent variable (churn) are categorical. The test statistic was 718.59, and the p-value was 9.1 x 10-157 (Figure 1). Because the p-value is less than 0.05, we reject the null hypothesis that customer churn is independent of contract type (Bruce and Gedeck, 2020; Griffiths, 2008). These results indicate that contract type and customer churn are likely related (Bruce and Gedeck, 2020; Griffiths, 2008).

**E2. Limitations of Analysis**

A statistically significant chi-square test tells us that two variables are related to one another, but it does not mean that there is a cause-effect relationship between the two variables. More statistical analysis is required to determine causality. Second, chi-square tests are sensitive to sample size, meaning that large samples can result in a statistically significant result even if the relationship isn’t strong.

**E3. Recommended Course of Action**

Based on the chi-square test, there is a relationship between contract type and customer churn. More data exploration and statistical analysis should be done to understand this relationship and determine causality before making any significant business decisions. If causality is established, and customers with longer contracts are truly less likely to churn, then the company should try to increase customers with one- and two-year contracts. To do this, the company could offer promotions or pricing discounts for customers who choose to sign up for a longer contract. In addition, they could offer a reward system for customers who re-up their contract after the initial one- to two- year time frame.

**F. Video**

A Panopto video was attached and submitted with the performance task. The link can also be accessed here: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=6326cdeb-0eaa-40c6-b538-afd2008b7da5>

1. **Sources of Third-Party Code**

No additional web sources of data or third-party code were used.

**H. Sources**

Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists : 50+ essential concepts using r and python*. O'Reilly Media, Incorporated.

Griffiths, D. (2008). *Head first statistics : A brain-friendly guide*. O'Reilly Media, Incorporated.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., … van Mulbregt,P. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.