Scenario: Telecommunications Churn

In the telecommunications industry, customers can choose from multiple service providers and actively switch from one provider to another. Customer "churn" is defined as the percentage of customers who stopped using a provider's product or service during a certain time frame. In this highly competitive market, some telecommunications industries can experience average annual churn rates as high as 25 percent. Given that it costs 10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.

For many providers, retaining highly profitable customers is the number one business goal. To reduce customer churn, telecommunications companies need to predict which customers are at high risk of churn.

You are an analyst on a team of analysts in a popular telecommunications company, which serves customers in all regions of the United States. You have been asked to clean the raw data set in preparation to explore the data, identify trends, and compare key metrics.

Data File Being Used:
churn_clean.csv

Data Dictionary:

The data set includes the following information:

- customers who left within the last month (the "Churn" column)
- services that each customer signed up for (phone, multiple lines, internet, online security, online backup, device protection, technical support, and streaming TV and movies)
- customer account information (how long they've been a customer, contracts, payment methods, paperless billing, monthly charges, GB usage over a year, etc.)
- customer demographics (gender, age, job, income, etc.)

The data set consists of 10,000 customers and 50 columns/variables:

- **CaseOrder:** A placeholder variable to preserve the original order of the raw data file
- **Customer_id:** Unique customer ID
- **Interaction, UID:** Unique IDs related to customer transactions, technical support, and sign-ups

   The following variables represent customer demographic data:
   - **City:** Customer city of residence as listed on the billing statement
   - **State:** Customer state of residence as listed on the billing statement
   - **County:** Customer county of residence as listed on the billing statement
   - **Zip:** Customer zip code of residence as listed on the billing statement
   - **Lat, Lng:** GPS coordinates of customer residence as listed on the billing statement
   - **Population:** Population within a mile radius of customer, based on census data
   - **Area:** Area type (rural, urban, suburban), based on census data
   - **TimeZone:** Time zone of customer residence based on customer's sign-up information
   - **Job:** Job of the customer (or invoiced person) as reported in sign-up information
   - **Children:** Number of children in customer's household as reported in sign-up information
   - **Age:** Age of customer as reported in sign-up information
   - **Education**: Highest degree earned by customer as reported in sign-up information
   - **Employment**: Employment status of customer as reported in sign-up information
   - **Income**: Annual income of customer as reported at time of sign-up
   - **Marital:** Marital status of customer as reported in sign-up information
   - **Gender:** Customer self-identification as male, female, or nonbinary
- **Churn:** Whether the customer discontinued service within the last month (yes, no)
- **Outage_sec_perweek:** Average number of seconds per week of system outages in the customer's neighborhood
- **Email:** Number of emails sent to the customer in the last year (marketing or correspondence)
- **Contacts:** Number of times customer contacted technical support
- **Yearly_equip_failure:** The number of times customer's equipment failed and had to be reset/replaced in the past year

- **Techie:** Whether the customer considers themselves technically inclined (based on customer questionnaire when they signed up for services) (yes, no)
- **Contract:** The contract term of the customer (month-to-month, one year, two year)
- **Port_modem:** Whether the customer has a portable modem (yes, no)
- **Tablet:** Whether the customer owns a tablet such as iPad, Surface, etc. (yes, no)
- **InternetService:** Customer's internet service provider (DSL, fiber optic, None)
- **Phone:** Whether the customer has a phone service (yes, no)
- **Multiple:** Whether the customer has multiple lines (yes, no)
- **OnlineSecurity:** Whether the customer has an online security add-on (yes, no)
- **OnlineBackup:** Whether the customer has an online backup add-on (yes, no)
- **DeviceProtection:** Whether the customer has device protection add-on (yes, no)
- **TechSupport:** Whether the customer has a technical support add-on (yes, no)
- **StreamingTV:** Whether the customer has streaming TV (yes, no)
- **StreamingMovies:** Whether the customer has streaming movies (yes, no)
- **PaperlessBilling:** Whether the customer has paperless billing (yes, no)
- **PaymentMethod:** The customer's payment method (electronic check, mailed check, bank (automatic bank transfer), credit card (automatic))
- **Tenure**: Number of months the customer has stayed with the provider
- **MonthlyCharge:** The amount charged to the customer monthly. This value reflects an average per customer.
- **Bandwidth_GB_Year:** The average amount of data used, in GB, in a year by the customer

The following variables represent responses to an eight-question survey asking customers to rate the importance of various factors/surfaces on a scale of 1 to 8 (1 = most important, 8 = least important)
  - **Item1:** Timely response
  - **Item2:** Timely fixes
  - **Item3:** Timely replacements
  - **Item4:** Reliability
  - **Item5:** Options
  - **Item6:** Respectful response
  - **Item7:** Courteous exchange
  - **Item8:** Evidence of active listening