Objective:

* This is an academic exercise using Customer Churn dataset intending to find answers to with the following problem statements
  + Is there a relationship between demographics and churn? Is particular type of customers more prone to churn over others? What role do partner, family size and dependents play w.r.t customer loyalty/churn rate?
  + Is there a relationship of bundled services with customer loyalty? What are the bundled service usage buckets/pattern? Is there any error/conflict potential in bundling (ex – availing tape-drive and cloud backup for live content)?
  + To reduce churn, what effort should I prioritize (if the choices or outcome of prior analysis) – customer service and interactions/follow ups, targeted marketing and sales campaigns, improving service quality and terms with specific identified areas? What is the impact of efforts/churn on customers count, revenue, valuation, etc?

##### Dataset: Customer Churn Data “WA\_Fn-UseC\_-Telco-Customer-Churn.csv”

Source: <https://www.kaggle.com/blastchar/telco-customer-churn> (IBM Sample Datasets]

Details:

* Each row represents a customer.
* Customers who left within last month – indicated by column Churn
* Dataset contains
  + Services availed/signed up by customer
  + Customer account/contract details
  + Customer Demographics

Data Exploration Observations:

* Churn Rates:
  + overall churn is 27%
  + gender doesn’t impact churn, females (27%) leaving is almost same as males (26%)
  + senior citizens (42%) have higher churn than non-seniors (24%)
  + relationship status does impact churn single churn is higher (33%) than customers with a partner (20%)
  + customers with dependents have lower (15%) churn than customers with dependents (31%)
  + single customers with no dependents have churn 34% -- they are probably college graduates
* Demographics
  + population 7043
  + genders are almost even females 49.5% and males 50.5%
  + senior citizens 16% of customers with 1142 customers. 476 churn indicates 7% churn w.r.t total customers
  + customers with partner are 48%, so churn customers are 9.5% of total
  + single customers 52% of total population, churn customers in these demographics make 17% of total
  + customers with dependents are 30% of total population, from this group churn customers make 5% of total
  + single customers without dependents are 47% of total population. Churn customers from this group make 16% of total
* Charges
  + avg monthly charge USD 64.76 (total USD 456, 116)
  + avg monthly charges for churn customers (USD 74) is higher than loyal customers (USD 61) i.e., customers who did not churn
  + senior citizens have higher monthly charges USD 80 (USD 79 for non-churn, USD 81 for churn customers) than non-senior citizens USD 62 (non-churn USD 59, churn USD 72)
  + customers with partner have marginally higher charges than single customers
  + customers with dependents have low monthly charges USD 59.5 (non-churn USD 57, churn USD 73)
  + single customers with dependents have even lower monthly charges USD 52.5 (non-churn USD 49, churn USD 65)
* Tenure
  + most customers leave during initial months, churn rates decrease thereafter and then again even out about 30 months
  + avg tenure overall is 32 months, with non-churn customers at 37.5 months and churn customers at 18 months
  + senior churn is higher in dataset, though a thing to note is senior avg tenure who churn is 21 months, and who do not is 42 months, it probably indicates the recency effect with timing when this dataset was captured
  + customers with partners are loyal customers 42 months (no-churn 45. Churn 26.5). single customers avg 23 months overall (churn customer 13 months)
* Contract Type
  + most popular contract is month-to-month i.e., 55% of total contracts. Two-year contracts (24%) and one-year (21%) are in that order
  + gender attribute has no impact on contract type
  + seniors have monthly contract 71% of cases while non seniors choose this popular contract 52%
  + customer with partners prefer monthly contracts, and also sign up for 2-year contract 35% of time
* Payment Methods
  + most customers pay with e-checks (38%). Other methods mailed check (21%), credit card (20%) and bank transfer (20.5%) are proportionate with respect to churn / non-churn customers (within 1% of each other proportions
  + gender doesn’t impact the choice with payment methods
* Paperless Billing
  + most customers prefer paperless billing (59%), with churn customers (75%) choosing it more than non-churn customers (54%)
  + seniors chose paperless billing is higher than overall (77%) and non-seniors are nearly similar as overall (56%)
  + relationship status doesn’t make any difference in overall preference for paperless billing
* Services – Internet, Phone, Streaming
  + fiber optic is the most popular one (internet service offering). DSL is more popular than fiber optic amongst non-churn clients. fiber optic is hands down the preferred service amongst churn customers
  + relationship status makes little difference with preference with internet service offerings
  + seniors chose fiber optics, and non-seniors DSL; DSL is also a clear winner amongst customers with dependents, customers without dependents opt for fiber optic
  + 21% of customers don’t use internet services.
  + more customers use phone service than internet offerings. Only 10% opt out. 1 line is most popular. multiple lines is higher for churn customers
  + customers with partners choose multiples at higher monthly rates, and customers with dependents chose single line contracts at slightly higher rate
  + 50% of customers signed up for streaming services. churn customers opted for streaming higher overall
  + seniors are using streaming at higher rate and 35% opt out of streaming services
  + customers with partners chose streaming services higher than average, while single customers opt out; customers with dependents opt out of streaming services at higher rates

Implications/ Conclusion:

1. I could build a customer persona of churn customer as
   1. not senior
   2. has partner
   3. has dependents
   4. no internet services
   5. has online backup, security, device protection, tech support devices
   6. 2-year contract
   7. Paper billing
   8. Low monthly charges
2. Such analysis/ persona outcomes are specifically relevant for Telecom firms who can now prioritize service offerings, marketing spends, and customer service interactions and focus on these targeted personas.

Afterthoughts:

1. I believe this analysis to provide directional bias, than exact inferences. The catch being with data accuracy and integrity
2. This study was conducted with academic assignment in purview, in real world, we would like to confirm the understanding with customer behavior by performing similar experiment on an alternate sourced dataset (i.e., maybe from a different company, for demographic indicator we shouldn’t be observing much difference, if the data were true representation of reality)
3. My goal was to use statistical concepts using R tools learnt in this course and apply these on a dataset to derive meaningful insight. I do believe I have accomplished this task objective, though I could have done better by including regression techniques and machine learning basics (say clustering techniques).

Reference:

1. <https://www.kaggle.com/farazrahman/telco-customer-churn-logisticregression>
2. <https://www.kaggle.com/blastchar/telco-customer-churn>